PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : H01L 29/788, 21/336

A1

(11) International Publication Number: WO 00/51188

(43) International Publication Date: 31 August 2000 (31.08.00)

(21) International Application Number: PCT/US00/04455

(22) International Filing Date: 17 February 2000 (17.02.00)

(30) Priority Data:

09/255,360	23 February 1999 (23.02.99)	US
09/275,670	24 March 1999 (24.03.99)	US
09/310,460	12 May 1999 (12.05.99)	US

(81) Designated States: CN, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

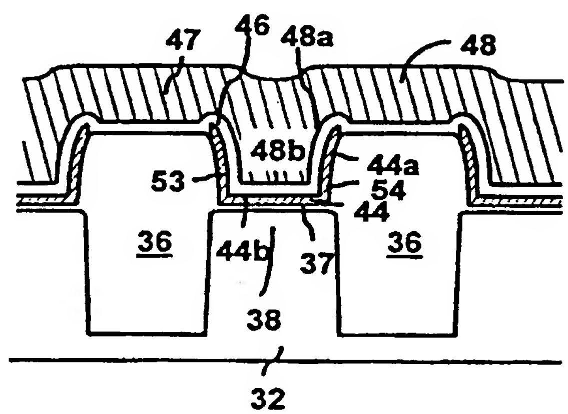
(71)(72) Applicant and Inventor: CHEN, Chiou-Feng [-/US]; 7573 Bollinger Road, Cupertino, CA 95014 (US).

(74) Agents: WRIGHT, Edward, S. et al.; Flehr Hohbach Test Albritton & Herbert LLP, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111-4187 (US).

(54) Title: FLASH MEMORY CELL WITH SELF-ALIGNED GATES AND FABRICATION PROCESS

(57) Abstract

Nonvolatile memory cell and process in which isolation oxide regions are formed on opposite sides of an active area in a substrate to a height above the substrate on the order of 80 to 160 percent of the width of the active area, a first layer of silicon is deposited on the gate oxide and along the sides of the isolation oxide regions to form a floating gate having a bottom wall which is substantially coextensive with the gate oxide and side walls having a height on the order of 80 to 160 percent of the width of the bottom wall, a dielectric film is formed on the floating gate, and a second layer of silicon is deposited on the dielectric film and patterned to form a control gate.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

IAP5 Rec'd PCT/PTO 27 MAR 2006

FLASH MEMORY CELL WITH SELF-ALIGNED GATES AND FABRICATION PROCESS

This invention pertains generally to semiconductor devices and, more particularly, to a flash memory cell with self-aligned gates and to a process for fabricating the same.

Electrically programmable read only memory (EPROM) has been widely used as nonvolatile memory which can keep data unchanged even though the power is turned off. However, EPROM devices have a major disadvantage in that they have to be exposed to Ultra-Violet (UV) light for about 20 minutes for data erasure. This is very inconvenient because an EPROM device has to be unplugged from its socket and moved to the UV light source when the data needs to be changed.

Electrically erasable programmable read only memory (EEPROM) overcomes this problem and permits data to be erased electrically in a much shorter period of time, typically less than 2 seconds. However, it still has a disadvantage in that the data must be erased on a byte-by-byte basis.

Flash EEPROM is similar to EEPROM in that data is erased electrically and relatively quickly. However, with flash EEPROM, the data is erased in blocks which typically range in size from 128 to 64K bytes per block, rather than on a byte-by-byte basis.

Examples of EPROM, EEPROM or flash EEPROM memory devices fabricated with conventional shallow trench and LOCOS (local oxidation of silicon)

isolation techniques are illustrated in Figures 1 and 2. Each of these devices includes an array of memory cells 11, each of which has a floating gate 12 and a control gate 13. The floating gate is an isolated island of polycrystalline or amorphous silicon which is formed above a thin gate oxide 14 in an active area 16 where the source, drain and channel regions (not shown) are located. The control gate is positioned above the floating gate, and is usually made of heavily doped polysilicon or polycide. A dielectric film 17 is positioned between the two gates. Depending upon the application, the dielectric can be an ONO (oxide/nitride/oxide) film, a pure oxide film, or another combination of oxide and nitride.

The threshold voltage of a memory cell, as observed from control gate, is dependent upon the amount of electrons stored inside the floating gate. Most memory cells can store two-bit data, *e.g.* a logic "1" for the conduction state when the threshold voltage is low, and a logic "0" for the non-conduction state when threshold voltage is high. In some high density applications, multi-level cells can store more than two bits per cell, *e.g.* 4 bits, 8 bits, or more, per cell. This is accomplished by controlling the amount of electrons inside floating gate more precisely so that more than two different threshold voltages can be achieved.

Heretofore, in most devices of this type, the floating gate has been formed by thermally growing the gate or tunnel oxide 14 to a thickness on the order of 70 - 250 Å on the active areas 16 of a silicon substrate 18 between isolation oxide regions 19 which separate adjacent ones of the memory cells 11. A conductive layer 21 is then formed on the gate oxide, and the dielectric film 17 is formed on the conductive layer. The conductive layer is typically a chemical vapor deposited (CVD) polysilicon film or an amorphous silicon film having a thickness on the order of 1500 - 2500 Å, and is doped with phosphorus, arsenic or boron either by *in situ* doping or by ion implantation. The dielectric film is either pure oxide or a combination of oxide and nitride.

A photolithographic mask is formed on the dielectric film to define either a complete floating gate pattern in which all four sides of the floating gate are delineated or a partial floating gate pattern in which only the two sides along the edges of the active area are delineated. The unmasked portions of the dielectric film and the conductive layer are etched away anisotropically to form the floating gate pattern. A second conductive layer 22 is then formed over the dielectric layer, and a second photolithographic mask is formed on the second conductive layer to define the control gate pattern and to complete the definition of the floating gate pattern in cases where that pattern was only partly defined previously. The unmasked portions of the second conductive layer and the dielectric layer are then etched away anisotropically to complete the control gate pattern and to complete the floating gate pattern where that pattern has not been completed previously.

Top views of stack-gate and split-gate memory cell arrays made with the shallow trench or LOCOS techniques are illustrated in Figures 3 and 4. The floating gates 12 have end caps 23, 24 which extend over the isolation oxide regions 19. Control gates 13 overlie the floating gates and form a word lines. In the split-gate array, the control gates include portions 13a which are used as select gates, with no portion of the floating gates beneath them. Bit lines 25, which are typically made of metal and separated by isolation oxide 19, interconnect the drains of the memory cells in each column. Source lines 26, which typically consist of P+ or N+ diffusion layers in the silicon substrate, interconnect the sources of the memory cells in the same row.

End caps 23, 24 are required in order to provide tolerance for corner rounding and shifting of the floating gate patterns relative to the isolation oxide regions during the photolithographic masking step(s) in which the floating gates are formed. The corner rounding effect may make the edges 27 shorter after the photolithographic step(s), and the shift of the floating

gate relative to the active area may cause the edges 27 to move beyond the edges 28 of the active areas. Either or both of those effects can result in a floating gate which does not completely cover the active area, and that can create a leakage path which will cause the transistor to malfunction.

In extending over the isolation oxide regions, end caps 23, 24 also help to form large capacitance areas 29 between the control gate and the floating gate, which results in a large coupling ratio between the two gates. This large ratio is important in memory cells because more voltage can be coupled from the control gate to the floating gate during write and erase operations.

In order to prevent the situation where the floating gate only partially covers the active area due to process variations, it has been necessary to increase the layout tolerance by making the end caps wider. In addition, the spacing 30 between adjacent ones of the floating gates must be made wide enough to avoid shorting between the gates. These two requirements have resulted in increased memory cell size and higher die cost.

U.S. Patent 5,767,005 describes a process for fabricating a self-aligned floating gate which does not have caps overlying the isolation, or field, oxide. In this process, the floating gate is formed by depositing a conductive layer over the field oxide as well as in a recess formed in the field oxide over the active region of the substrate. The conductive layer is then planarized by a step such as chemical mechanical planarization (CMP) until the top of the conductive material is coplanar with the upper surface of the field oxide and the material is left only in the recess. The floating gate is thus self-aligned, with the inner wall of the recess defining the peripheral shape of the gate. While this process does eliminate the need for a critical masking step in the formation of the floating gate, it has the

disadvantage of applying CMP polishing directly to the polysilicon or other conductive material from which the floating gate is formed. In addition, the floating gate is relatively short and provides only limited coupling with the control gate.

It is in general an object of the invention to provide a new and improved semiconductor device and fabrication process.

Another object of the invention is to provide a semiconductor device and process of the above character which overcome the limitations and disadvantages of the prior art.

These and other objects are achieved in accordance with the invention by providing a nonvolatile memory cell and process in which isolation oxide regions are formed on opposite sides of an active area in a substrate to a height above the substrate on the order of 80 to 160 percent of the width of the active area, a gate oxide is formed over the active area, a first layer of silicon is deposited on the gate oxide and along the sides of the isolation oxide regions to form a floating gate having a bottom wall which is substantially coextensive with the gate oxide and side walls having a height on the order of 80 to 160 percent of the width of the bottom wall, a dielectric film is formed on the floating gate, and a second layer of silicon is deposited on the dielectric film and patterned to form a control gate which is capacitively coupled with the floating gate.

In some embodiments, a control gate or a thick dielectric film is used as a mask in the formation of a floating gate and also as a step in the formation and alignment of a select gate. The floating gate is relatively thin and has a side wall with a rounded curvature which, in some embodiments, serves as a tunneling window for electrons migrating to the select gate during erase operations. In other embodiments, the gate oxide beneath the

floating gate is relatively thin, and the electrons tunnel through the gate oxide to the source region in the substrate below.

Figure 1 is a schematic cross sectional view of a floating gate memory device of the prior art with shallow trench isolation.

Figure 2 is a schematic cross sectional view of a floating gate memory device of the prior art with LOCOS isolation.

Figure 3 is a schematic top plan view of a stack-gate memory cell array of the prior art.

Figure 4 is a schematic top plan views of a split-gate memory cell array of the prior art.

Figures 5A - 5H are schematic cross sectional views illustrating the steps in one embodiment of a process for fabricating a stack-gate or split-gate memory cell using shallow trench isolation in accordance with the invention.

Figure 6 is a schematic top plan of a stack-gate memory device manufactured in accordance with the process of Figures 5A - 5H.

Figure 7 is a cross-sectional view taken along line 7-7 in Figure 6.

Figures 8A - 8E are schematic cross sectional views illustrating the steps in another embodiment of a process for fabricating a stack-gate or split-gate memory cell using LOCOS isolation in accordance with the invention.

Figure 9 is a schematic top plan of a split-gate memory cell array manufactured in accordance with the process of Figures 8A - 8E.

Figure 10 is a cross-sectional view taken along line 10-10 in Figure 9.

Figures 11A and 11B are schematic cross sectional views illustrating the steps in another embodiment of a process for fabricating a nonvolatile memory device using shallow trench isolation in accordance with the invention.

Figures 12A and 12B are schematic cross sectional views illustrating the steps in another embodiment of a process for fabricating a nonvolatile memory device using LOCOS isolation in accordance with the invention.

Figures 13A - 13G are schematic cross sectional views illustrating the steps in one embodiment of a process for fabricating a flash memory cell with self-aligned floating, control, and select gates.

Figures 14 is a cross sectional view of an embodiment of a flash memory cell which has self-aligned floating and select gates, but no control gate.

Figure 15 is a cross sectional view of another embodiment of a flash memory cell which has self-aligned floating, control and select gates.

Figures 16 - 19 are top plan views of memory cell arrays utilizing the memory cells of Figures 13G, 14 and 15.

Figures 20 and 21 are circuit diagrams of the memory cell arrays of Figures 16 - 19.

As illustrated in Figure 5A, shallow trenches 31 are formed in a silicon substrate 32. The silicon can be an N-well material, a P-well material or simply a P-type material. A pad oxide 33 is formed on the substrate, and a silicon nitride layer 34 is deposited on the pad oxide. These layers are

patterned to form a mask, and the substrate is etched through the mask to form the trenches.

An isolation oxide 36 is then deposited in the trenches and planarized so that it is level with the upper surface of the nitride layer. The oxide can, for example, be deposited by chemical vapor deposition (CVD), and planarized by CMP polishing. In this process, pad oxide 33 and nitride layer 34 serve not only as a mask for forming the trenches, but also as a means for building up the height of the isolation oxide. As discussed more fully hereinafter, the step height 35 (*i.e.* the height of the upper surface of the nitride layer above the upper surface of the silicon substrate) is important because it provides the floating gate with a large side wall along the edge of the isolation oxide, which increases the capacitance between the control gate and the floating gate. With a large coupling ratio, more voltage is coupled from the control gate to the floating gate during write and erase operations of the memory cell.

After the isolation oxide regions are formed, the nitride layer is stripped away, and a thin silicon layer 39 is deposited as illustrated in Figure 5B. This layer is etched anisotropically to form silicon spacers along the edges of the isolation oxide, as illustrated in Figure 5C. These spacers are then thermally oxidized to form steeply sloped oxide spacers, as illustrated in Figure 5D. Thereafter, pad oxide 33 is removed from the silicon substrate by a wet dip, and a thin gate or tunnel oxide 37 is formed on active areas 38, as illustrated in Figure 5E. The gate oxide is thermally grown and has a thickness on the order of 70 - 150 Å.

A relatively thin silicon layer 41 is then deposited on the gate oxide and the isolation oxide, and a nitride film 42 is deposited on the silicon. The silicon layer has a thickness on the order of 100 - 1000 Å, which is substantially thinner than the floating gates (1500 - 2500 Å) of prior art devices. It can be either polysilicon or amorphous silicon. The silicon is preferably doped

with phosphorus, arsenic or boron to a level on the order of 10^{17} to 10^{20} per cm^3. The doping can be done *in-situ* during deposition of the silicon or by ion implantation through the nitride.

A silicon-on-glass (SOG) or CVD oxide 43 is deposited over the nitride film, and planarized by etch back or CMP polishing to the level of the upper surface of the nitride film over the isolation oxide 36, as illustrated in Figure 5F. That leaves oxide 43 only in the regions above active areas 38, and using that oxide as a mask, the nitride film is etched away in the areas over the isolation oxide. The remaining SOG or CVD oxide is then etched away with a wet chemical solution, exposing the nitride 42 in the active areas. Using that nitride as a mask, the silicon 41 above the isolation oxide is removed by an anisotropic dry etch, as shown in Figure 5G, leaving the silicon only in the active areas to form floating gates 44. Those gates extend along the top surface of gate oxide 37 and along the side edges of the isolation oxide.

As illustrated in Figure 5H, once the floating gates have been formed, the nitride which covers them is stripped away, and a dielectric film 46 is deposited over the floating gates and the isolation oxide. That film is preferably an ONO (oxide/nitride/oxide) film, but it can also be a pure oxide film or another combination of oxide and nitride. A layer 47 of polysilicon or polycide (*e.g.*, tungsten silicide) is deposited over the dielectric film and patterned to form the control gates 48, as illustrated in Figure 6.

The control gates are patterned by means of a photolithographic mask (not shown) which also defines the edges 51 of the floating gates which extend across active areas 38. The two silicon layers and the dielectric layer in the unmasked areas are then etched away anisotropically, leaving control gates 48 and floating gates 44, with edges 51 of the floating gates being self-aligned with the corresponding edges of the control gates. The other

two edges 52 of the floating gates are self-aligned with the side edges 36a of the isolation oxide.

As illustrated in Figure 7, drain and source regions 53, 54 are formed in substrate 18, with portions 54a of the source regions lying beneath the floating gates. These regions are heavily doped conduction layers with a conductivity type opposite to that of the silicon substrate. Oxide spacers 55 are formed along the sides of the memory cells.

The capacitance areas between the control gates and the floating gates can be made relatively large by making silicon layer 41 relatively thin and increasing the height of the side walls 44a of the floating gates. In the embodiment illustrated, the height of the side walls 44a is on the order of 80 to 160 percent of the width of the bottom wall 44b of the floating gate, and preferably is on the order of 1000 - 5000 Å. The control gates have relatively high side walls 48a and bottom walls 48b in closely spaced facing relationship with the side walls and bottom walls of the floating gates. This provides a large coupling ratio between the control gates and the floating gates so that voltages can be coupled efficiently from the control gates to the floating gates, and the memory cells can maintain a robust write and erase performance without floating gate caps over the isolation oxide.

Operation of the memory cell of Figures 6 and 7 is as follows, with bias voltages applied to the three node terminals as set forth in Table 1.

Table 1

Mode	Control Gate	Drain	Source
Erase (1)	0 volts	Floating	≈ 12 volts
Erase (2)	≈ - 7 volts	Floating	≈ 5 volts
Program	≈ 10 volts	≈ 5 volts	0 volts
Read	3 to 5 volts	1.5 to 3 volts	0 volts

In the erase mode, electrons inside floating gate are forced to tunnel from floating gate to the portions of the source regions 54a beneath the floating gates by Fowler-Nordheim tunneling mechanism. During an erase operation, a high electric field (> 10MV/cm) is established across the tunnel oxide 37. This can be achieved by applying a negative voltage of about -7 volts to the control gate and a positive voltage of about 5 volts to the source node, or by applying 0 volts to the control gate and about 12 volts to the source node. The drain node is kept floating in both cases. With a coupling ratio of about 90 percent between the control gate and the floating gate in the erase mode, most of the voltage difference between the source and the control gate appears across the tunnel oxide, with electrons being forced to tunnel from the floating gate to the overlapped portion of the source region.

When the erase operation is completed, the floating gate is positively charged, the threshold voltage of memory cell becomes lower, and the memory cell is in a conduction, or logic "1", state.

In the program mode, electrons are injected into the floating gate through hot carrier injection, and the floating gate becomes negatively charged. During a programming operation, the control gate is biased at about 10 volts, the drain is biased at about 5 volts, and the source is biased at 0 volts. When electrons flow from source to drain during programming, they

are accelerated by high electric field across the channel region, and some of them become heated near the drain junction. Some of the hot electrons can surpass the oxide barrier height of about 3.1 eV and inject into the floating gate.

When the programming operation is completed, the floating gate is negatively charged, the threshold voltage of memory cell becomes higher, and the memory cell is in non-conduction, or logic "0" state.

In the read mode, the control gate is biased at about 3 to 5 volts, the source is biased at 0 volts, and drain is biased at about 1.5 to 3 volts. When the memory cell is in the erase state, the read shows a conduction state, and a logic "1" is identified by the sense amplifier. When memory cell is in the program state, the read shows a non-conduction state, and a logic "0" is identified by the sense amplifier.

The process illustrated in Figures 8A - 8E is generally similar to the process of Figures 5A - 5H except that it uses LOCOS (local oxidation of silicon) isolation rather than shallow trench isolation. Like reference numerals designate corresponding elements in the devices made by the two processes.

In this embodiment, after the pad oxide 33 is formed on the substrate 32, a polysilicon layer 56 is formed on the pad oxide, and a nitride layer 57 is formed on the polysilicon layer. The isolation oxide 36 is thermally grown, with the shortest possible bird's beak 58 projecting laterally from the oxide. The height 59 of the isolation oxide above the substrate is important in providing the floating gate with a high side wall and a large coupling capacitance between the control gate and the floating gate. In the embodiment illustrated, the height of the isolation oxide above the substrate is on the order of 55 percent of the total height 61 of the oxide.

After the isolation oxide regions are formed, the pad oxide, polysilicon and nitride are stripped away, and the gate oxide 37 is thermally grown on the active areas 38. Thereafter, the silicon layer 41, nitride layer 42 and SOG or CVD oxide 43 are deposited as in the embodiment of Figures 5A - 5H, and the SOG or CVD oxide is planarized to a level even with the upper surface of the nitride over the isolation oxide, thus leaving the SOG or CVD oxide 43 only in the active areas.

As in the embodiment of Figures 5A - 5H, the nitride above the isolation oxide 36 is etched away, using the SOG or CVD oxide as a mask. The SOG or CVD oxide is then removed from the active areas with a wet etching solution, and using the remaining nitride 42 as a mask, the silicon 41 above the isolation oxide is etched away, leaving the silicon only in the active regions.

The nitride 42 covering the silicon in the active regions is stripped away, and a photolithographic mask is employed to define one or both of the sides 51, 52 of the floating gates 44 which extend across the active regions 38. If only one side is defined, it is the side 51 which faces toward contact 66, and the other side 52 is delineated during formation of the control gate pattern. After the floating gate mask is defined, the silicon in the unmasked areas is etched away anisotropically.

The dielectric film 46 is deposited over the remaining silicon and the isolation oxide, and the second silicon layer 47 is deposited on the dielectric film. That layer can be formed of polysilicon which preferably is heavily doped with phosphorus, arsenic or boron to a level on the order of 10^{20} to 10^{21} per cm^3. The doping can be done either by *in-situ* doping during deposition or by ion implantation. Alternatively, a polycide such as tungsten silicide can be used instead of polysilicon.

After the second silicon layer is formed, another photolithographic mask is employed to define the control gate pattern, as illustrated in Figure 9. Following pattern definition, the unmasked portions of the layer are etched away anisotropically to form the control gates 48. Where the second sides 52 of the floating gates have not already been formed, the anisotropic etching continues through the dielectric film 46 and silicon layer 44 to complete the floating gate pattern.

As illustrated in Figure 10, a portion 48c of control gate 48 overhangs floating gate 44, a relatively thin gate oxide 37a (*e.g.,* 70 - 150 Å) is formed between the silicon substrate and the floating gate, and a relatively thick gate oxide 37b (*e.g.,* 100 - 350 Å) is formed between the substrate and the overhanging portion of the control gate. Channel 66 also has one portion 66a beneath the overhanging portion of the control gate and another portion 66b beneath the floating gate. In this embodiment, drain 53, channel 66 and control gate 48 can be considered to be a select transistor which is directly adjacent to the memory cell of which floating gate 44 and control gate 48 are a part.

As in the device manufactured by the process of Figures 5A - 5H, the floating gates have relatively high side walls 44a which, in the embodiment illustrated, are on the order of 80 to 160 percent of the width of the bottom wall 44b, and are preferably on the order of 1000 - 5000 Å high. The control gates also have relatively high side walls 48a and bottom walls 48b in closely spaced facing relationship with the side walls and bottom walls of the floating gates. This provides a relatively large area for capacitive coupling with the control gates and results in a large coupling ratio so that voltages can be coupled efficiently from the control gates to the floating gates. As a result, the memory cells can maintain a robust write and erase performance without floating gate caps over the isolation oxide.

Operation of the memory cell of Figures 9 and 10 is as follows, with bias voltages applied to the three node terminals as set forth in Table 2.

Table 2

Mode	Control Gate	Drain	Source
Erase (1)	0 volts	Floating	≈ 12 volts
Erase (2)	≈ - 7 volts	Floating	≈ 5 volts
Program	≈ 10 volts	0 volts	≈ 7 volts
Read	3 to 5 volts	1.5 to 3 volts	0 volts

In the erase mode, electrons inside floating gate are forced to tunnel from floating gate to the portions of the source regions 54a beneath the floating gates by Fowler-Nordheim tunneling mechanism. During an erase operation, a high electric field (> 10MV/cm) is established across the tunnel oxide 37. This can be achieved by applying a negative voltage of about -7 volts to the control gate and a positive voltage of about 5 volts to the source node, or by applying 0 volts to the control gate and about 12 volts to the source node. The drain node is kept floating in both cases. With a coupling ratio of about 90 percent between the control gate and the floating gate in the erase mode, most of the voltage difference between the source and the control gate appears across the tunnel oxide, with electrons being forced to tunnel from the floating gate to the overlapped portion of the source region.

When the erase operation is completed, the floating gate is positively charged, the threshold voltage of memory cell becomes lower, and the memory cell is in a conduction, or logic "1", state.

In the program mode, electrons are injected into the floating gate through hot carrier injection, and the floating gate becomes negatively charged.

During a programming operation, the control gate is biased at about 10 volts, the drain is biased at 0 volts, and the source is biased at about 7 volts. When electrons flow from drain to source during programming, they are accelerated by high electric field across the channel region, and some of them become heated near the source junction. Some of the hot electrons can surpass the oxide barrier height of about 3.1 eV and inject into the floating gate.

When the programming operation is completed, the floating gate is negatively charged, the threshold voltage of memory cell becomes higher, and the memory cell is in non-conduction, or logic "0" state.

In the read mode, the control gate is biased at about 3 to 5 volts, the source is biased at 0 volts, and drain is biased at about 1.5 to 3 volts. When the memory cell is in the erase state, the read shows a conduction state, and a logic "1" is identified by the sense amplifier. When memory cell is in the program state, the read shows a non-conduction state, and a logic "0" is identified by the sense amplifier.

Figures 11A and 11B illustrate another process for forming a self-aligned floating gate using CMP polishing and shallow trench isolation. Isolation oxide regions 36 are formed as described above, and gate oxide 37 is grown on the active regions 38. A layer 68 of polysilicon or amorphous silicon is deposited over the gate oxide and the isolation oxide. This layer has a thickness on the order of 100 - 1000 Å, and is doped with phosphorus, arsenic or boron to a level on the order of 10^{17} to 10^{20} per cm^3. A nitride layer 69 which also has a thickness on the order of 100 - 1000 Å is deposited on the silicon.

In the case where the silicon is very thin, *e.g.* less than about 500Å, and is doped by ion implantation, it is preferable to implant the ions through the

nitride so that most of the implanted ions are distributed into the silicon, rather than penetrating into the gate oxide and/or the silicon substrate.

A CMP polishing operation is then performed to remove the nitride 69 above the isolation oxide to a level at or slightly below the upper surface of the silicon 68 on the isolation oxide, as illustrated in Figure 11B. Using the nitride remaining in the active areas as a mask, the silicon above the isolation oxide is anisotropically etched away. The control gates, the dielectric film between the control gates and the floating gates, and the remaining edges of the floating gates are formed as in the process of Figures 5A - 5H.

Figures 12A and 12B illustrate a process which is substantially identical to the process of Figures 11A and 11B, except that the isolation oxide regions are formed by a LOCOS process, rather than shallow trenching. Once the isolation oxide regions are formed, the remaining steps are substantially identical, and like reference numerals designate corresponding elements in the devices produced by the two processes. Thus, gate oxide 37 is grown on the substrate, silicon layer 68 is deposited over the gate oxide and the isolation oxide, and nitride layer 69 is deposited on the silicon.

The nitride above the isolation oxide is removed by CMP polishing, and the remaining nitride is used as a mask in etching away the silicon above the isolation oxide. The control gates, the dielectric film between the control gates and the floating gates, and the remaining edges of the floating gates are formed as in the process of Figures 5A - 5H.

The invention has a number of important features and advantages. The relatively thin floating gate with high side walls provides a large coupling capacitance between the control gate and the floating gate, which provides robust write and erase operation without the end caps required by prior art

devices. Eliminating the end caps significantly reduces the size of the memory cells and the array. Moreover, process yield fluctuation caused by pattern shifting and corner-rounding are eliminated, resulting in better and more stable process yields.

The processes of Figures 11A - 11B and 12A - 12B have the further advantage that the thin silicon on the isolation oxide regions can be etched entirely during the silicon etch by adding an over-etch step, and at the same time the height 71 of the side wall of the silicon can be accurately controlled. This is important in maintaining a large capacitance between the control gate and the floating gate to get the desired amount of coupling. Moreover, the nitride also serves to protect the thin silicon layer and the underlying thin gate oxide from contamination during the CMP polishing step.

In the process illustrated in Figures 13A - 13G, an oxide layer 136 having a thickness on the order of 70 - 250 Å is thermally grown on a monocrystalline silicon substrate 137 for use as a gate oxide or a tunnel oxide. The substrate can be in the form of a P-well or P-substrate material. A layer 138 of polysilicon or amorphous silicon (the poly-1 layer) is deposited on the oxide layer for use as a floating gate, and a dielectric film 139 is formed on the silicon layer. The poly-1 layer has a thickness on the order of 100 - 1000 Å, which is substantially thinner than the floating gates in prior art devices which are typically about 2000Å thick. The poly-1 layer is preferably doped with phosphorus, arsenic or boron to a level on the order of 10^{17} to 10^{20} per cm^3. The doping can be done *in-situ* during deposition of the silicon or by ion implantation directly into the silicon or through the dielectric film.

The dielectric film can be either a pure oxide or a combination of oxide, nitride and oxide (ONO), and in the embodiment illustrated, it consists of a lower oxide layer 141 having a thickness on the order of 30 - 100 Å, a

central nitride layer 142 having a thickness on the order of 60 - 300 Å, and an upper oxide layer 143 having a thickness on the order of 30 - 100 Å.

A layer 144 of polysilicon (the poly-2 layer) is deposited on the dielectric film for use as a control gate. This layer has a thickness on the order of 1500 - 3500 Å, and is doped with phosphorous, arsenic or boron to a level on the order of 10^{20} to 10^{21} per cm^3. A layer 146 of CVD oxide or nitride is then deposited on the poly-2 layer to a thickness on the order of 300 - 1000 Å. During subsequent dry etching steps, the layer of oxide or nitride serves as a mask to prevent the poly-2 in the control gate area from being etched away.

A photolithographic mask (not shown) is formed over layer 146 to simultaneously define both the control gate and the floating gate, and the unmasked portions of that layer and the poly-2 layer are removed in an anisotropic dry etch, leaving only the portion of the poly-2 which forms the control gate 147, as illustrated in Figure 13B. The photoresist is then stripped away, and an oxide layer 148 is thermally grown on the side wall of the polysilicon to a thickness on the order of 100 - 600 Å.

During the thermal oxidation process, the corners of the polysilicon become rounded because the oxidation rate for the silicon is faster next to the interfaces with oxide layers 143, 146. This rounding enhances the performance of the memory cell during erase cycles by eliminating the sharp corners of the polysilicon and thus minimizing the leakage current between the control gate and the select gate.

Referring now to Figure 13C, using the control gate and oxide and/or nitride on it as a mask, the inter-poly dielectric and the poly-1 are etched in an anisotropic dry etch to form the inter-poly dielectric 149 and the floating gate 151. Thereafter, in a thermal oxidation step, a select gate oxide 152 if formed on the substrate, an oxide layer 153 is formed on the

exposed edge of the floating gate, and the oxide layer 148 on the side wall of the control gate is made thicker. Oxide layer 153 is used as a tunneling oxide, and preferably has a thickness on the order of 50 - 300 Å. If desired, a thin layer of CVD oxide (about 50 - 200 Å) can be deposited on the thermal oxide to improve the quality of the oxide films and reduce disturbances between the select gate and the floating gate.

During the thermal oxidation process, the edge portion or side wall 154 of the floating gate becomes rounded because the oxidation rate for the poly-1 is faster near the interfaces with the oxide layers above and below it. This rounded curvature provides an electric field enhancement which makes it much easier for electrons to tunnel out of the floating gate through this area. In addition, it eliminates the localized trapping effect which occurs in a tunnel oxide near square corners of poly-1. This enhances performance of the memory cell during both the program cycle and the erase cycle.

A polysilicon layer (poly-3) 156 is then deposited over the oxide layers, and a nitride or oxide layer 157 is deposited on the poly-3 layer, as illustrated in Figure 13D. During this deposition, the step formed by the control gate, the layer of oxide or nitride on it, the inter-poly dielectric and the floating gate causes a corresponding step 156a to be formed in the poly-3 layer. That step is utilized in subsequent removal of portions of the poly-3 layer to form the select gate, and for that reason, the thickness of the poly-3 layer should be less than the height of the step over which it is formed. In one present embodiment, the poly-3 layer has a thickness on the order of 2000 - 4000 Å, and nitride or oxide layer 157 has a thickness on the order of 200 - 1500 Å.

The nitride or oxide is removed from the flat areas of the poly-3 layer by anisotropic dry etching, leaving only the vertically extending portion 158, as shown in Figure 13E. The poly-3 layer is also etched anisotropically to

form the select gate 159. Since the poly-3 layer is thinner than the step formed by the control gate and the other elements beneath it, etching to the thickness of the poly-3 layer removes all of the poly-3 above the control gate and all of the poly-3 outside the area where the step 156a was formed.

The nitride or oxide in region 158 protects the shoulder 161 of the poly-3 from being etched away. It also provides control over the channel length of the select gate and reduces the sheet resistance of the select gate. Thus, for example, by making the nitride or oxide layer thicker, the shoulder can be made wider, resulting in a longer channel length. The thickness of the poly-3 layer can also be used to adjust the width and height of the select gate, with thicker poly-3 resulting in a thicker and wider gate. Lower sheet resistance reduces the loading effect of the select gate and results in faster performance for the memory cell.

The remaining nitride or oxide is stripped away, as illustrated in Figure 13F, and after photolithographic masking, the poly-3 spacer 162 on the side of the control gate opposite the select gate is etched away.

As illustrated in Figure 13G, source 163 and drain 164 are now formed in the substrate, and oxide spacers 166 are formed to the side of the select gate 161 and control gate 147. Both the source and the drain are heavily doped with an N-type material such as phosphorus or arsenic, and the source is made with a deeper junction than the drain in view of the high voltage which will be applied across it during the program operation. The erase path of this memory cell is from the rounded curvature of the side wall 154 of the floating gate to the select gate. Because of the relatively thin floating gate and the rounded curvature of the side wall, the coupling ratio between the control gate and the floating gate can be made large, and electron tunneling is more efficient due to enhancement of the local electric field around the curvature of the side wall.

By using the control gate as a mask, the floating gate is automatically self-aligned with the control gate. The select gate is likewise completely self-aligned with respect to the floating gate and the control gate, and the select gate can have a small sheet resistance and better performance during cell operation.

Operation of the memory cell of Figure 13G is as follow, with bias voltages applied to the four node terminals as set forth in Table 3.

Table 3

Mode	Control Gate	Select Gate	Drain	Source
Erase	-7.0 to -12.0	3.0 to 7.0	Floating	Floating
Program	7.0 to 10.0	1.5 to 3.0	0	4.0 to 8.0
Read	0 to 2.5	2.5 to 5.0	1.5 to 3.0	0

In the erase mode, Fowler-Nordheim tunneling causes electrons to migrate from the floating gate to the select gate, leaving the positive ions as the majority carrier inside the floating gate. The rounded curvature of the floating gate enhances the electric field effect, and the tunneling takes place at a lower applied voltage than it would with two flat surfaces on opposite sides of the tunneling oxide. The electric field enhancement also makes it possible to form a thicker tunnel oxide while still maintaining sufficient electron tunneling.

The coupling ratio from the control gate to the floating gate is about 70 to 80 percent in the erase mode, which means that about 70 to 80 percent of the control gate voltage is coupled to the floating gate. This causes most of the voltage difference between the control gate and the select gate to fall across the tunnel oxide surrounding the rounded side wall of the floating gate, which triggers Fowler-Nordheim tunneling with electrons tunneling from the floating gate to the select gate. As the floating gate

becomes more positive, the threshold voltage of the memory cell decreases to a level of -5.0 to -1.0 volts. This results in an inversion layer of the channel under the floating gate when the control gate is biased at 0 to 2.5 volts. Therefore, the memory cell is in a conduction, or logic 1, state after an erasing.

In the program mode, electrons are injected into the floating gate, and the floating gate becomes negatively charged. With the control gate biased at 7.0 to 10 volts, the select gate biased at 1.5 to 3.0 volts, the drain biased at around 0 volts, and the source biased at 4.0 to 8.0 volts, most of the source-to-drain voltage falls across the mid-channel region between the select gate and the floating gate, which generates a high electric field in the mid-channel region. Moreover, since the floating gate is coupled to a high voltage by the voltages supplied from the source and the control gate, a high vertical electric field is established across the oxide between the mid-channel region and the floating gate.

When electrons flow from the drain to the source during programming, they are accelerated by the high field across the mid-channel region, and some of them become heated. Being accelerated by the vertical electric field, some of the hot electrons can surpass the oxide barrier height of about 3.1 eV and inject into the floating gate. When the programming is finished, the floating gate becomes negatively charged, and the threshold voltage of the memory cell increases to a level of about 3.0 to 6.0 volts. This turns off the channel under the floating gate, with the control gate biased at 0 to 2.5 volts. Thus, the memory cell is in a non-conductive, or logic 0, state after programming.

In the read mode, the control gate is biased at 0 to 2.5 volts, the source is biased at 0 volts, the drain is biased at 1.5 to 3.0 volts, and the select gate is biased at 2.5 to 5.0 volts. When the memory cell is in the erase state, the read shows a conduction state (logic 1) because both the

floating gate and the select gate channels are turned on. When the memory cell is in a program state, the read shows a non-conduction state (logic 0) because the floating gate channel is turned off.

The embodiment of Figure 14 is generally similar to the embodiment of Figure 13G, and like reference numerals designate corresponding elements in the two embodiments. In the embodiment of Figure 14, however, there are only two silicon layers, and the memory cell has only a floating gate 151 and a select gate 159; there is no control gate. This embodiment also has a deeper source junction 167 and a wider area of source overlap 168 under the floating gate, which serve the function of the control gate in providing the voltage which is coupled to the floating gate.

The ONO film on the poly-1 is also thicker in this embodiment, with a bottom oxide layer 169 which is on the order of 50 - 300Å thick, a nitride layer 171 which is on the order of 1000 - 2000Å thick, and a top oxide layer 172 which is on the order of 200 - 1000Å thick. This thick film is etched anisotropically to serve as a mask in the formation of the floating gate and to provide a step which is used in the formation of the select gate, just as the control gate is utilized in the embodiment of Figures 13A-4G. Alternatively, if desired, the ONO layer can be replaced with an ON film having a bottom oxide layer on the order of 50 - 300Å thick and a nitride layer on the order of 1000 - 3000Å thick. The erase path of this memory cell is from the rounded curvature of the side wall 154 of the floating gate to the select gate.

Operation of the memory cell of Figure 14 is as follows, with bias voltages applied to the node terminals as set forth in Table 4a.

Table 4a

Mode	Select Gate	Drain	Source
Erase	12 to 15.0	Floating	0
Program	1.5 to 3.0	0	10.0 to 13.0
Read	2.5 to 5.0	1.5 to 3.0	0

In the erase mode, Fowler-Nordheim tunneling causes electrons to tunnel from the floating gate to the select gate. A positive voltage on the order of 12 to 15 volts is applied to the select gate, the drain node is kept floating, and the source node is biased at 0 volts. With a coupling ratio from the source node on the order of 70 to 85 percent, for example, about 70 to 85 percent of the source voltage is coupled to the floating gate. With the combination of the overlapping source 168 and channel regions 173, there is more voltage coupled from the source node to the floating gate. This results in a larger coupling ratio, *e.g.*, about 80 to 90 percent in this example. Therefore, most of the voltage difference between the select gate and the source is located across the tunnel oxide surrounding the rounded side wall 154 of the floating gate, triggering Fowler-Nordheim tunneling and forcing electrons to tunnel from the floating gate to the select gate. After the erase operation is finished, the floating gate becomes positively charged, and the threshold voltage of the memory cell decreases to the level of -5.0 to -1 volt. Thus, an inversion layer can be formed in the channel under the floating gate even though there is no control gate above the floating gate. The memory cell is now in its conduction state (logic 1).

In the program mode, the select gate is biased at 1.5 to 3.0 volts, the drain is biased at around 0 volts, and the source is biased at 10 to 13 volts. This bias condition can turn on the channel under the select gate and the channel under the floating gate. Therefore, most of the source-to-drain voltage is set across the mid-channel region between the select gate

and the floating gate. The floating gate is coupled to a high voltage from the source under the floating gate, and thus a high vertical electric field is established across the oxide between the mid-channel region and the floating gate. The combined high electric fields around the mid-channel region generates hot electrons and forces some of them to be injected into floating gate. After programming is finished, the floating gate becomes negatively charged, and the threshold voltage of the memory cell increases to a level on the order of 3.0 to 6.0 volts. The channel under the floating gate is now turned off, and the memory cell is in a nonconductive state (logic 0).

In the read mode, the source is biased at 0 volts, the drain is biased at 1.5 to 3.0 volts, and the select gate is biased at 2.5 to 5.0 volts. When the memory cell is in the erase state, the read shows a conduction state because both the floating gate and the select gate channels are turned on. When the memory cell is in the program state, the read shows a nonconductive state because the floating gate channel is turned off.

Alternatively, in the embodiment of Figure 14, the memory cell can be fabricated in a P-well surrounded by an N-well or an N-type substrate. In this case, the bias voltages applied to the terminal nodes of the memory cell are set forth in Table 4b.

Table 4b

Mode	Select Gate	Drain	Source	P-Well
Erase	6.0 to 9.0	Floating	-9 to -6	-9 to -6
Program	1.5 to 3.0	0	10.0 to 13.0	Floating
Read	2.5 to 5.0	1.5 to 3.0	0	Floating

Operation of this memory cell is otherwise the same as described above for the embodiment of Figure 14, with the channel and source regions beneath

the floating gate providing an even higher coupling ratio for the floating gate.

The embodiment of Figure 15 is also generally similar to the embodiment of Figure 13G, and like reference numerals designate corresponding elements in these two embodiments as well. In the embodiment of Figure 15, however, oxide layer 136 is on the order of 70 - 120Å thick, and is used as the tunnel oxide for the erase operation. The oxide 153 surrounding the rounded side wall 154 of the floating gate is not used as a tunneling oxide in this embodiment, and can be made thicker, *e.g.*, up to about 200 - 1000Å. A poly-2 layer 144 is included to form a control gate 147, and the inter- poly dielectric is preferably an ONO film, as in the embodiment of Figure 13G. In this embodiment, the tunneling window is the area of overlap 168 of the source beneath the floating gate, and the erase path is from the floating gate to the overlapping portion of the source.

Operation of the memory cell of Figure 15 is as follows, with bias voltages applied to the four node terminals as set forth in Table 5.

Table 5

Mode	Control Gate	Select Gate	Drain	Source
Erase	-7.0 to -12.0	Floating	Floating	3.0 to 7.0
Program	7.0 to 10.0	1.5 to 3.0	0	4.0 to 8.0
Read	0 to 2.5	2.5 to 5.0	1.5 to 3.0	0

In the erase mode, electrons inside the floating gate are forced to tunnel from the floating gate to the overlapping source region 168 by Fowler-Nordheim tunneling. During the erase operation, a high electric field (> 10MV/cm) is established across the tunnel oxide 136 that was thermally grown on the silicon substrate. That is achieved by applying a negative

voltage on the order of -7.0 to -12.0 to the control gate and a positive voltage on the order of 3.0 to 7.0 volts to the source node. The drain and select gate nodes are kept floating, and the coupling ratio can, for example, be on the order of 70 to 85 percent. Therefore, most of the voltage difference between the control gate and the source is located across the tunnel oxide, resulting in Fowler-Nordheim tunneling and movement of electrons from the floating gate to the overlapping source region. After the erasing operation is finished, the floating gate is positively charged, and the threshold voltage of the memory cell drops to a level on the order of -5.0 to -1.0 volts. This forms an inversion layer in the channel under the floating gate when the control gate is biased at 0 to 2.5 volts. Therefore, the memory cell is in a conduction state (logic 1) after an erase operation.

The program mode is similar to that of the embodiment of Figure 13G. However, the values of the bias voltages may be vary due to the different thickness of oxide 136 and oxide 153. After the programming is finished, the floating gate becomes negatively charged, and the threshold voltage of the memory cell increases to a level on the order of 3.0 to 6.0 volts. This turns off the channel under the floating gate when the control gate is biased to 0 to 2.5 volts. Therefore, the memory cell is in a nonconducting state (logic 0) after programming.

Figures 16 and 17 illustrate NOR-type memory cell arrays of memory cells of the type illustrated in Figures 13G and 15 with and without floating gates which are self-aligned with the edges of the active areas. In the embodiment of Figure 16, the edges 151a, 151b of floating gate 151 are aligned with the edges of active area of substrate 137, whereas in the embodiment of Figure 17, the floating gate has end caps 151c, 151d which extend over isolation oxide regions 173. Without self-alignment of the edges of the floating gate with the edges of the active region, these end caps are necessary in order to prevent short channeling or

punchthrough due to pattern shifting or corner rounding during formation of the floating gates.

A circuit diagram for the memory cell arrays of Figures 16 and 17 is shown in Figure 20. All of the memory cells in a given column have their drains connected to a bit line BL_{n-1}, BL_n, BL_{n+1}, etc., which is typically a metal line 176-78 that crosses over the active area, and all of the cells in a given row are connected to a source line 179, which is typically an N+ diffusion layer in the silicon substrate 137. Adjacent ones of the bit lines are isolated from each by a dielectric film (not shown). All of the select gates 159 in a given row are connected together by a select gate line 181, and all of the control gates 147 in a given row are connected together by a control gate line 182. The select gate lines and the control gate lines are formed of the poly-3 and poly-2 layers, respectively.

Figures 18 and 19 illustrate NOR-type memory cell arrays of memory cells of the type illustrated in Figure 14 with and without floating gates which are self-aligned with the edges of the active areas. In the embodiment of Figure 18, the edges 151a, 151b of floating gate 151 are aligned with the edges of active area of substrate 137, whereas in the embodiment of Figure 19, the floating gate has end caps 151c, 151d which extend over isolation oxide regions 173. These embodiments are similar to the embodiments of Figures 16 and 17 except there are no control gates.

A circuit diagram for the memory cell arrays of Figures 18 and 19 is shown in Figure 21. This circuit is similar to the circuit of Figure 20 except there are no control gates.

The invention has a number of important features and advantages. The floating gate and the select gate are self-aligned both with respect to each other and with respect to the control gate. Being relatively thin, the floating gate can have a larger coupling ratio with the control gate, and it

presents a smaller tunneling window when the side wall or edge of the gate is used as a source of electrons during erase operations. Moreover, the rounded side wall or edge of the floating gate enhances the local electric field around the curvature of the edge and provides more efficient tunneling of electrons during erase operations. This permits the tunnel oxide layer between the select gate and the floating gate to be made wider but still maintains a robust erase operation.

Since the select gate is formed from a layer of silicon which is deposited over a step created by the control gate or a thick dielectric layer, the height and width of the select gate can be controlled by making the step higher or shorter, by depositing a thicker layer of silicon, and by depositing a thicker or thinner layer or nitride or oxide above the silicon. The nitride or oxide protects the shoulder of the silicon step during etching to form the select gate. Consequently, the select gate can have a small sheet resistance, a small loading effect, and faster performance.

It is apparent from the foregoing that a new and improved semiconductor device and process have been provided. While only certain presently preferred embodiments have been described in detail, as will be apparent to those familiar with the art, certain changes and modifications can be made without departing from the scope of the invention as defined by the following claims.

CLAIMS

1. In a nonvolatile memory cell: a substrate having an active area, a gate oxide on the active area, isolation oxide extending upwardly from the substrate on opposite sides of the gate oxide to a height on the order of 80 to 160 percent of the width of the gate oxide, a floating gate above the gate oxide having a thin bottom wall and side walls which extend to the top of the isolation oxide so that the height of the side walls is on the order of 80 to 160 percent of the width of the bottom wall, a dielectric film overlying the floating gate, and a control gate above the dielectric film coupled capacitively to the floating gate.

2. The memory cell of Claim 1 wherein the bottom wall of the floating gate has a thickness on the order of 100 to 1000 Å.

3. The memory cell of Claim 1 wherein the floating gate is fabricated a silicon material selected from the group consisting of polysilicon and amorphous silicon.

4. The memory cell of Claim 3 wherein the silicon material is doped with a material selected from the group consisting of phosphorus, arsenic and boron.

5. The memory cell of Claim 1 wherein the gate oxide has a thickness on the order of 70 to 250 Å.

6. The memory cell of Claim 1 wherein a side edge of the floating gate is aligned with a side edge of the control gate.

7. The memory cell of Claim 1 wherein two side edges of the floating gate are aligned with two side edges of the control gate.

8. In a process for fabricating a semiconductor device having a floating gate, the steps of: forming isolation oxide regions on opposite sides of an active area in a substrate to a height above the substrate on the order of 80 to 160 percent of the width of the active area, forming a gate oxide layer over the active area, depositing a first layer of silicon on the gate oxide and along the sides of the isolation oxide regions to form a floating gate having a bottom wall which is substantially coextensive with the gate oxide and side walls having a height on the order of 80 to 160 percent of the width of the bottom wall, forming a dielectric film on the floating gate, and depositing a second layer of silicon on the dielectric film to form a control gate which is capacitively coupled with the floating gate.

9. The process of Claim 8 wherein the isolation oxide regions are formed by forming a temporary layer on the substrate to a height on the order of 80 to 160 percent of the width of the active area, depositing oxide to a level above the temporary layer, planarizing the oxide to the height of the temporary layers, and removing the temporary layer.

10. The process of Claim 9 wherein a portion of the oxide is deposited in shallow trenches in the substrate.

11. The process of Claim 8 wherein the isolation oxide regions are formed by forming a temporary layer on the substrate to a height on the order of 80 to 160 percent of the width of the active area, growing oxide to a level above the temporary layer, planarizing the oxide to the height of the temporary layers, and removing the temporary layer.

12. In a process for fabricating a semiconductor device having a floating gate, the steps of: forming an isolation oxide on a substrate on opposite sides of an active region, forming a gate oxide on the substrate in the active region, depositing a first layer of silicon on the gate oxide and on the sides and top of the isolation oxide, forming a layer of nitride on the first

layer of silicon, forming a temporary oxide on the nitride layer, planarizing the temporary oxide to the top of the nitride to expose the nitride above the isolation oxide and leaving the temporary oxide in the active region, etching away the nitride above the isolation oxide using the temporary oxide as a mask and leaving the nitride in the active region, removing the temporary oxide from the nitride in the active region, etching away the silicon above the isolation oxide using the nitride in the active region as a mask and leaving the silicon in the active region to form a floating gate which is aligned with the sides of the isolation oxide, removing the nitride from the active region to expose the floating gate, forming a dielectric film on the floating gate and on the isolation oxide, depositing a second layer of silicon on the dielectric film, and patterning the second layer of silicon to form a control gate which is coupled capacitively with the floating gate through the dielectric film.

13. The process of Claim 12 including the steps of positioning a mask above the second silicon layer, and etching the control gate and the floating gate through the mask to form aligned side edges on the two gates.

14. The process of Claim 12 including the steps of forming one side edge on the floating gate by etching through a first mask before the second layer of silicon is formed, and etching through a second mask during the patterning of the second layer to form another side edge on the floating gate.

15. In a process for fabricating a semiconductor device having a floating gate, the steps of: forming isolation regions on a substrate on opposite sides of an active region, forming a gate oxide on the substrate in the active region, depositing a first layer of silicon on the gate oxide and on the sides and top of the isolation region, forming a layer of nitride on the first layer of silicon, planarizing the nitride to the level of the silicon above the

isolation regions, etching away the silicon above the isolation regions using the nitride in the active region as a mask and leaving the silicon in the active region to form a floating gate, removing the nitride from the floating gate, forming a dielectric film over the floating gate and the isolation regions, forming a second layer of silicon over the dielectric film, and patterning the second layer to form a control gate.

16. The process of Claim 15 including the step of doping the first layer of silicon with a dopant selected from the group consisting of phosphorus, arsenic and boron.

17. The process of Claim 16 wherein the first layer of silicon is doped by implanting ions through the nitride layer.

18. In a process for fabricating a semiconductor device, the steps of: forming a floating gate in an active region above a substrate, forming a control gate above the floating gate, using a photolithographic mask to define aligned side edges on the control gate and the floating gate, and etching through the mask to form the aligned side edges.

19. The process of Claim 18 including the step of using another photolithographic mask to define another side edge of the floating mask prior to formation of the control gate.

20. In a process for fabricating a semiconductor device, the steps of: forming a floating gate in an active region between two isolation regions, using a first photolithographic mask to define two side edges of the floating gate perpendicular to the isolation regions, etching through the first mask to form the side edges, forming a control gate above the floating gate, using a second photolithographic mask to define two side edges of the control gate, and etching through the second mask to form the two side edges of the control gate.

21. In a memory cell: a substrate having an active area, an oxide layer formed on the substrate above the active area, a relatively thin floating gate having a side wall with a rounded curvature positioned above the oxide layer, a control gate which is substantially thicker than the floating gate positioned above and in vertical alignment with the floating gate, a dielectric film between the floating gate and the control gate, a select gate positioned to one side of the control gate and facing the side wall of the floating gate with the rounded curvature, a tunnel oxide between the select gate and the floating gate, and a tunneling path for the migration of electrons during erase operations extending from the side wall of the floating gate with the rounded curvature through the tunnel oxide to the select gate.

22. The memory cell of Claim 21 wherein the floating gate has a thickness on the order of 100 - 1000Å, and the control gate has a thickness on the order of 1500 - 3500Å.

23. In a memory cell: a substrate having an active area, an oxide layer formed on the substrate above the active area, a relatively thin floating gate having a side wall with a rounded curvature positioned above the oxide layer, a dielectric film which is substantially thicker than the floating gate positioned above and in vertical alignment with the floating gate, a select gate positioned to one side of the control gate and facing the side wall of the floating gate with the rounded curvature, a tunnel oxide between the select gate and the floating gate, and a tunneling path for the migration of electrons during erase operations extending from the side wall of the floating gate with the rounded curvature through the tunnel oxide to the select gate.

24. The memory cell of Claim 23 wherein the dielectric film comprises a bottom oxide layer which is on the order of 50 - 300Å thick, a nitride

layer which is on the order of 1000 - 2000Å thick, and a top oxide layer which is on the order of 200 - 1000Å thick.

25. The memory cell of Claim 23 wherein the dielectric film comprises an oxide layer which is on the order of 50 - 300Å thick and a nitride layer which is on the order of 1000 - 3000Å thick.

26. In a memory cell: a substrate having an active area, a relatively thin gate oxide layer formed on the substrate above the active area, a relatively thin floating gate having a side wall with a rounded curvature positioned above the oxide layer, a control gate which is substantially thicker than the floating gate positioned above and in vertical alignment with the floating gate, a dielectric film between the floating gate and the control gate, a select gate positioned to one side of the control gate and facing the side wall of the floating gate with the rounded curvature, a relatively thick oxide layer between the select gate and the floating gate, a source region formed in the substrate with a portion of the source region overlapping beneath the floating gate, and a tunneling path for the migration of electrons during erase operations extending from the floating gate through the gate oxide layer to the overlapping portion of the source region.

27. The memory cell of Claim 26 wherein the gate oxide layer has a thickness on the order of 70 - 120Å, and the oxide layer between the select gate and the floating gate has a thickness on the order of 200 - 1000Å.

28. In a process of fabricating a semiconductor device having a floating gate, a control gate and a select gate, the steps of: forming an oxide layer in an active area on a silicon substrate; forming a first silicon layer on the oxide layer; forming a dielectric film on the first silicon layer; forming a second silicon layer on the dielectric film; etching away a portion of the second silicon layer to form a control gate; using the control gate as a

mask, anisotropically etching away portions of the dielectric film and the first silicon layer to form a floating gate beneath the control gate; forming a third silicon layer over the substrate and the control gate with a step in the third silicon layer beside and above the control gate; and anisotropically etching the third silicon layer to form a select gate beside the control gate.

29. The process of Claim 28 wherein the thickness of the third silicon layer is less than the total thickness of the layers beneath it where it steps over the control gate.

30. The process of Claim 28 further including the step of forming a rounded curvature on a side wall of the floating gate.

31. The process of Claim 30 wherein the rounded curvature is formed by oxidation of the side wall, with the oxidation occurring at a greater rate near the oxide layer beneath the floating gate and near the dielectric film above it.

32. The process of Claim 28 wherein the dielectric film is formed by forming a bottom oxide layer on the first silicon layer, forming a nitride layer on the bottom oxide layer, and forming a top oxide layer on the nitride layer.

33. The process of Claim 28 further including the steps of forming a nitride layer on the third silicon layer, and anisotropically removing the nitride everywhere except in a shoulder area of the step in the third silicon layer.

34. The process of Claim 28 further including the steps of forming an additional oxide layer on the third silicon layer, and anisotropically removing the additional layer everywhere except in a shoulder area of the step in the third silicon layer.

35. The process of Claim 28 wherein the oxide layer is formed with a thickness on the order of 70 - 250Å, the first silicon layer is formed with a thickness on the order of 100 - 1000Å, the dielectric film is formed with a thickness on the order of 120 - 500Å, the second silicon layer is formed with a thickness on the order of 1500 - 3500Å, and the third silicon layer is formed with a thickness on the order 2000 - 4000Å.

36. The process of Claim 28 including the steps of doping each of the silicon layers with a dopant selected from the group consisting of phosphorus, arsenic and boron.

37. The process of Claim 28 wherein the first silicon layer is doped to a level on the order of 10^{17} to 10^{20} per cm^3.

38. The process of Claim 28 wherein the second and third silicon layers are doped to a level on the order of 10^{20} to 10^{3} per cm^3.

39. In a process of fabricating a semiconductor device, the steps of: forming an oxide layer in an active area on a silicon substrate; forming a first silicon layer on the oxide layer; forming a relatively thick dielectric film above the first silicon layer; anisotropically etching the relatively thick dielectric film to form a step above the active area; using the step as a mask, anisotropically etching the first silicon layer to form a floating gate above the active area; forming a second silicon layer over the substrate and the step; and anisotropically etching the second silicon layer to form a select gate beside the floating gate.

40. The process of Claim 39 wherein the dielectric film is formed by forming a bottom oxide layer to a thickness on the order of 50 - 300Å thick, forming a nitride layer on the bottom oxide layer to a thickness on

the order of 1000 - 2000Å thick, and forming a top oxide layer on the nitride layer to a thickness on the order of 200 - 1000Å.

41. The process of Claim 39 wherein the dielectric film is formed by forming a bottom oxide layer on the first silicon layer to a thickness on the order of 50 - 300Å, and forming a nitride layer on the bottom oxide layer to a thickness on the order of 1000 - 3000Å.

42. The process of Claim 39 wherein the thickness of the second silicon layer is less than the height of the step.

43. The process of Claim 39 further including the step of forming a rounded curvature on a side wall of the floating gate.

44. The process of Claim 43 wherein the rounded curvature is formed by oxidation of the side wall, with the oxidation occurring at a greater rate near the oxide layer beneath the floating gate and near the dielectric film above it.

Fig. 1 (Prior Art)

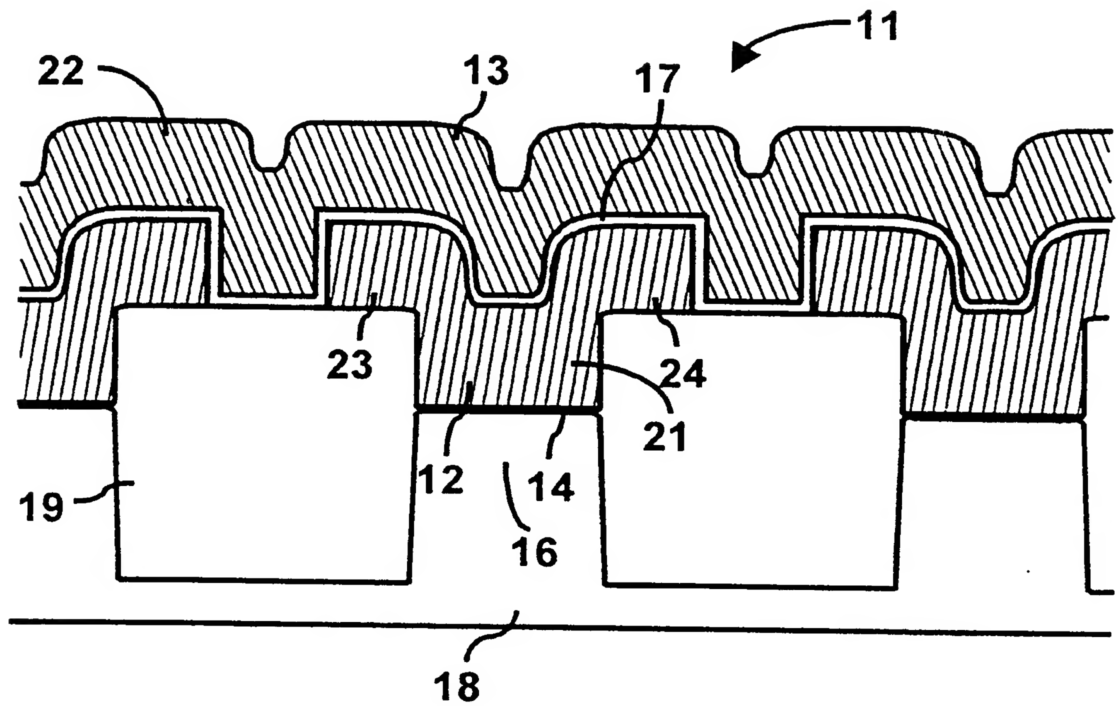


Fig. 2 (Prior Art)

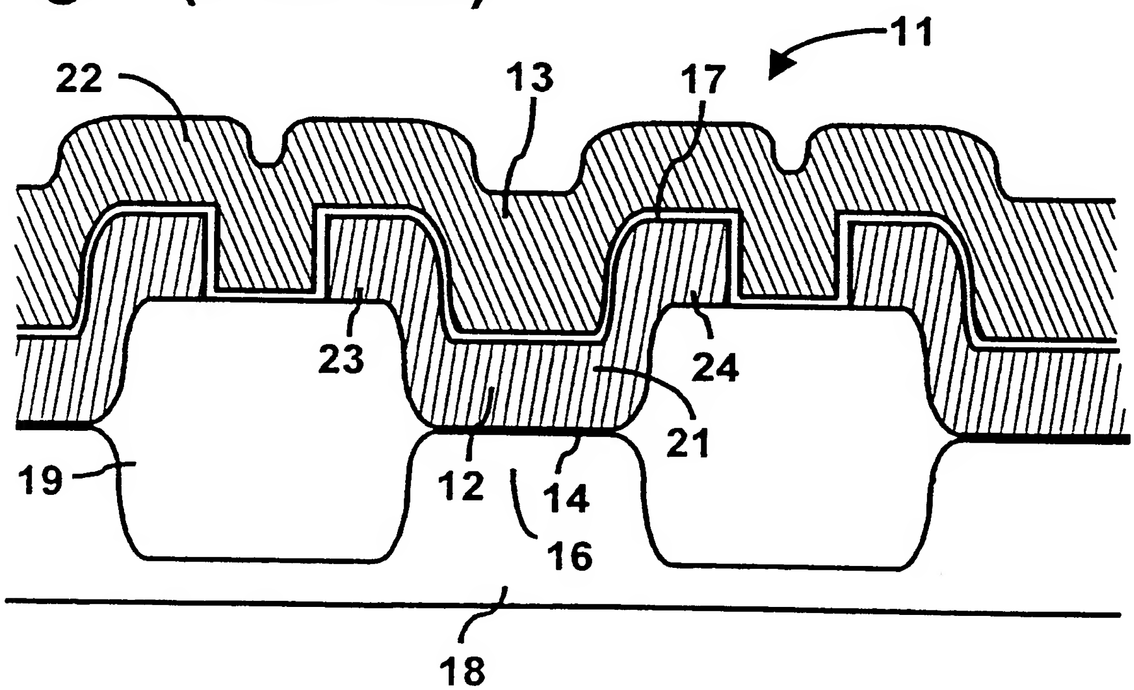


Fig. 3 (Prior Art)

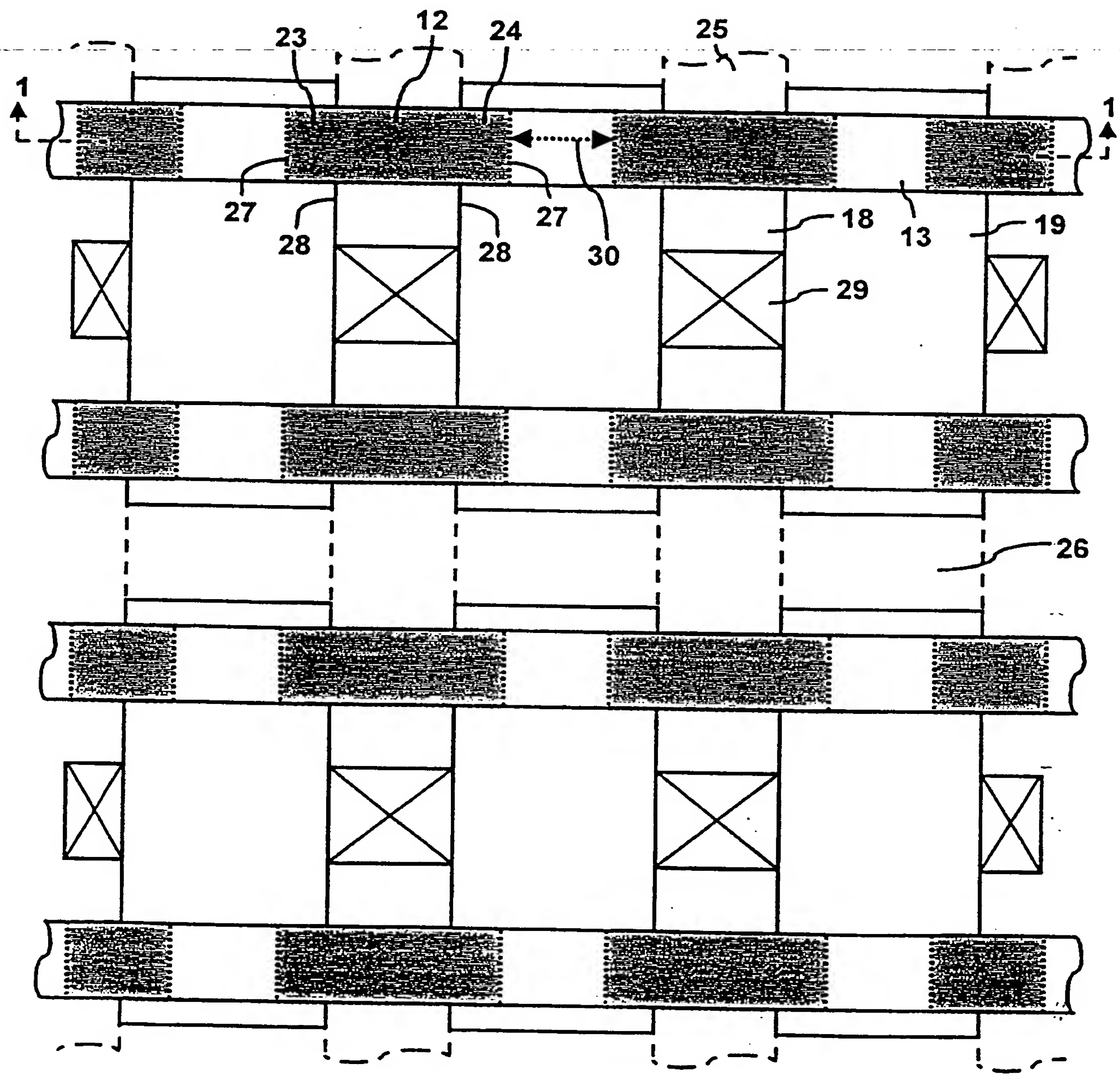


Fig. 4 (Prior Art)

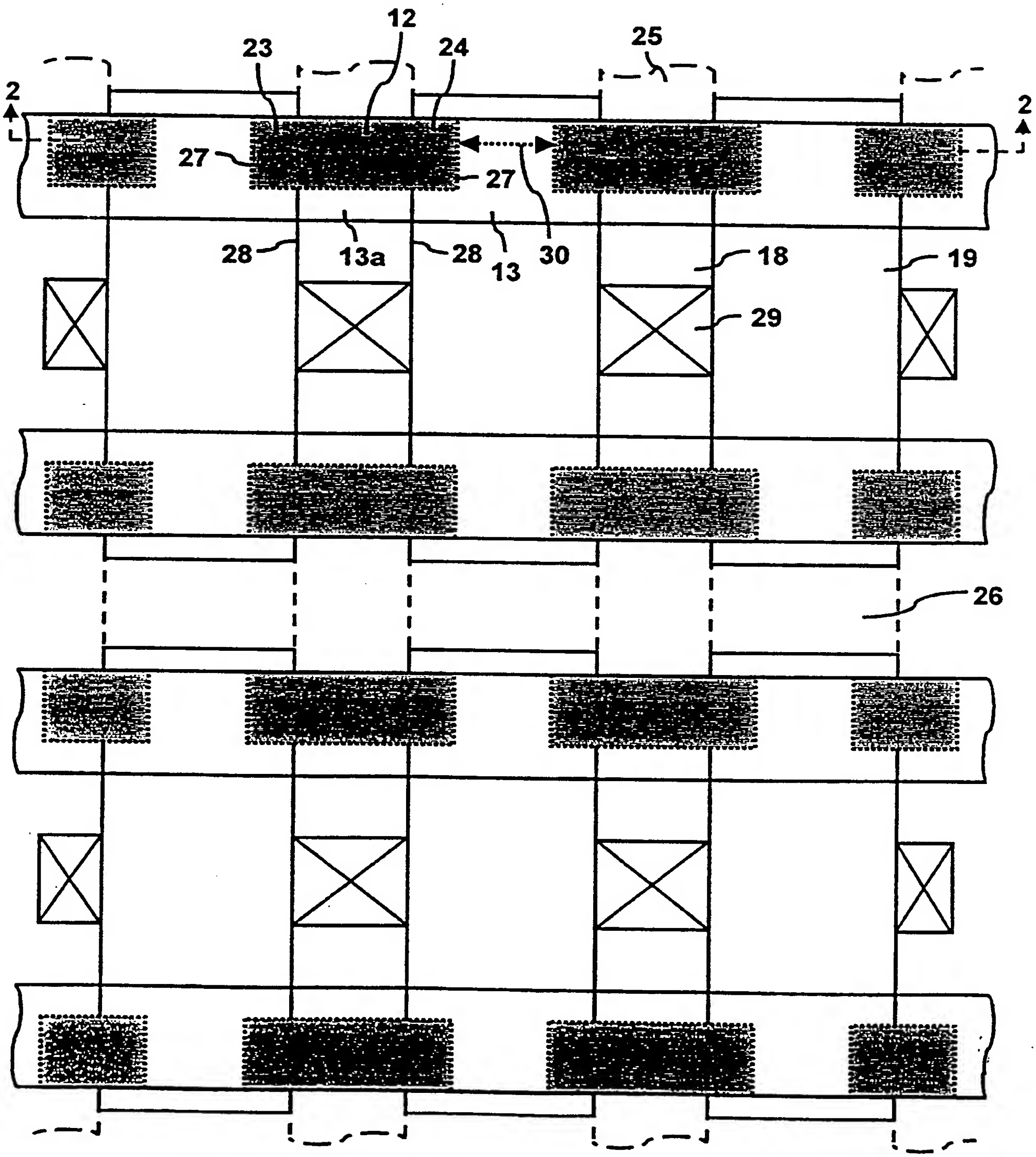


Fig. 5A

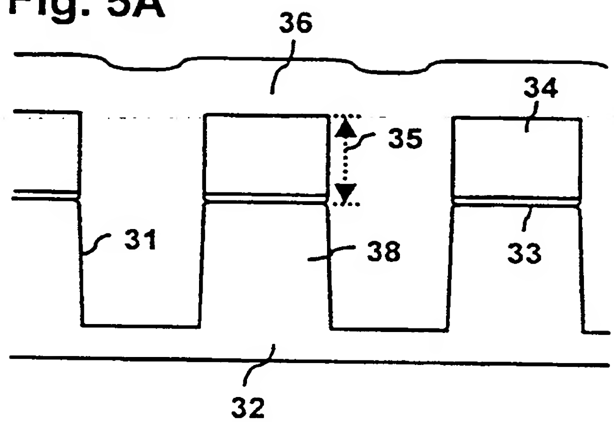


Fig. 5B

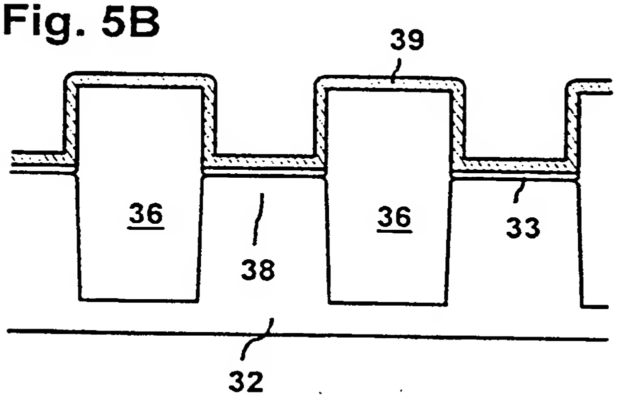


Fig. 5C

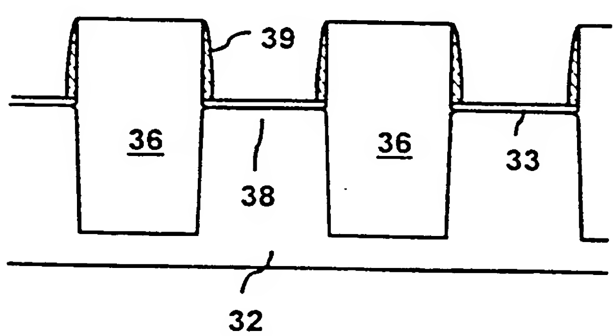


Fig. 5D

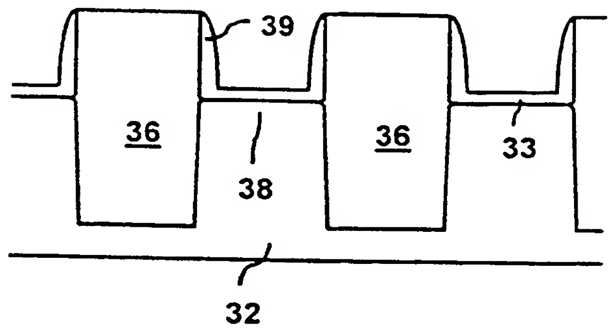


Fig. 5E

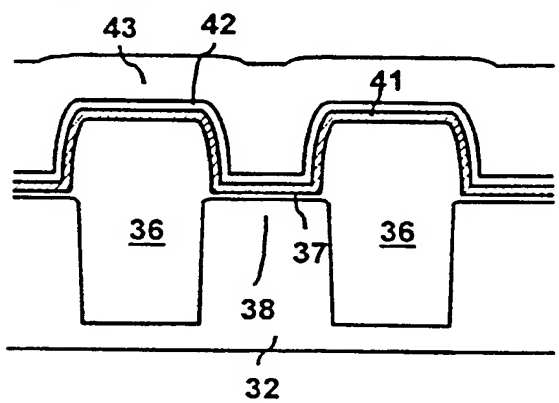


Fig. 5F

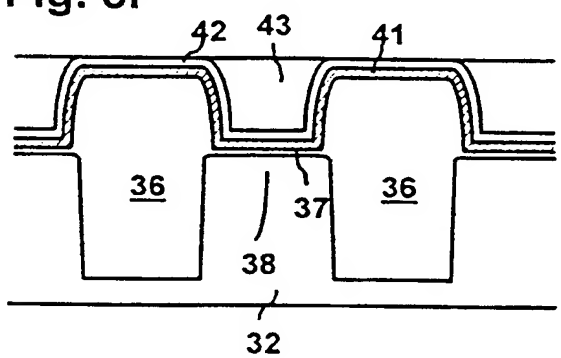


Fig. 5G

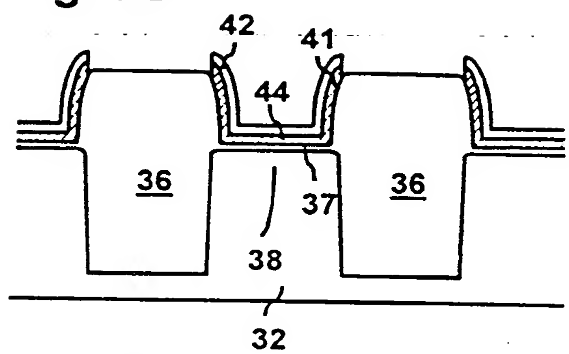

42
41
44
36
37
36
38
32

Fig. 5H

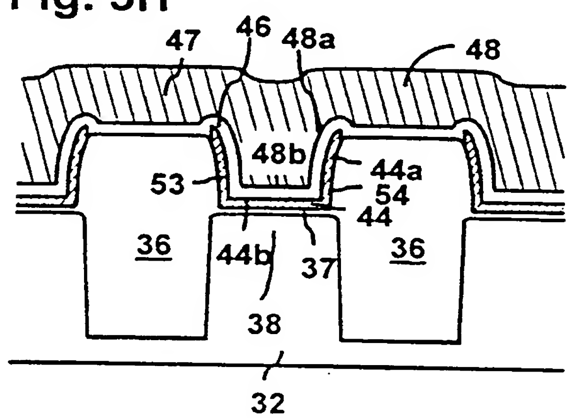
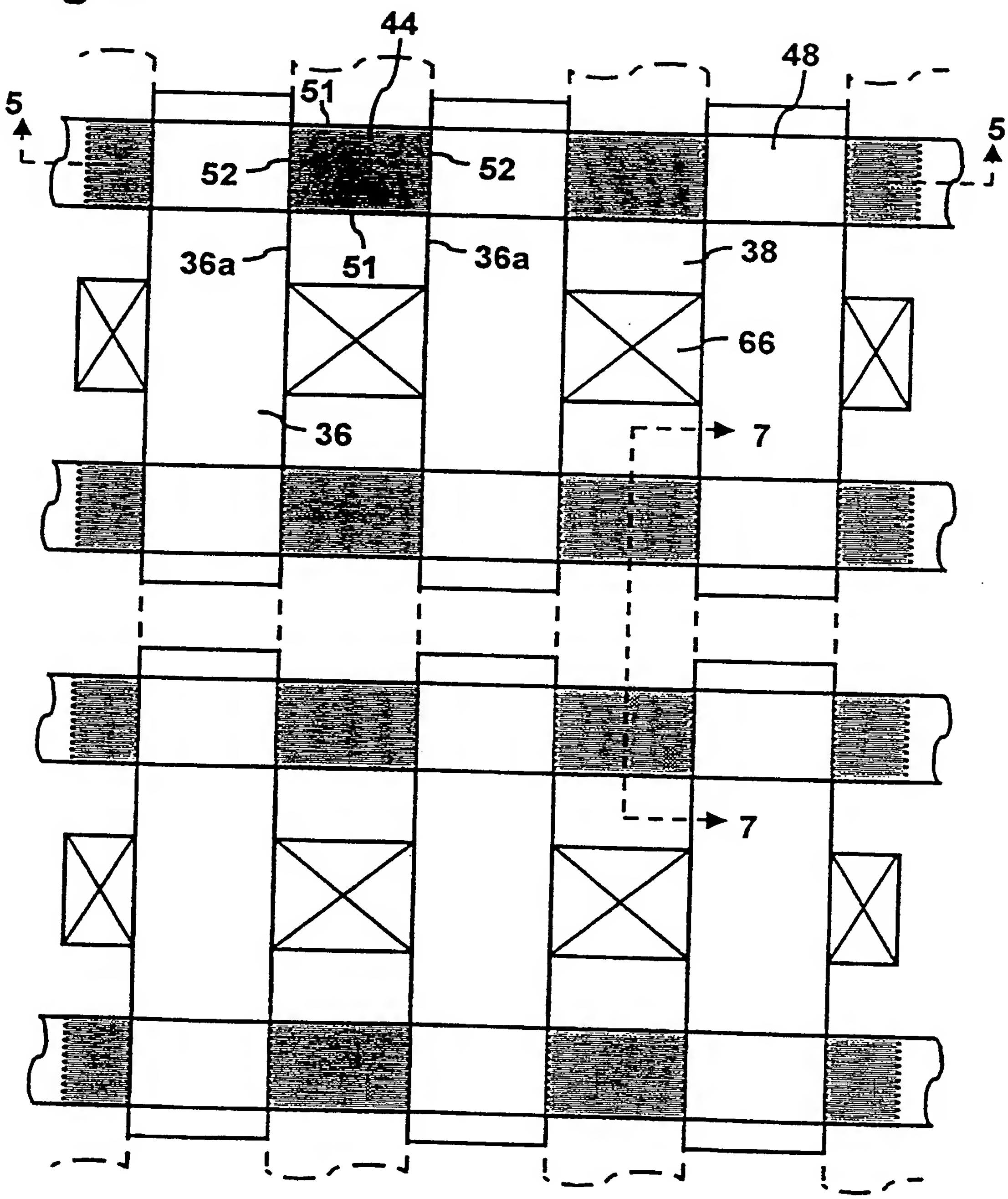

47
46
48a
48
48b
44a
53
54
44
36
44b
37
36
38
32

Fig. 6



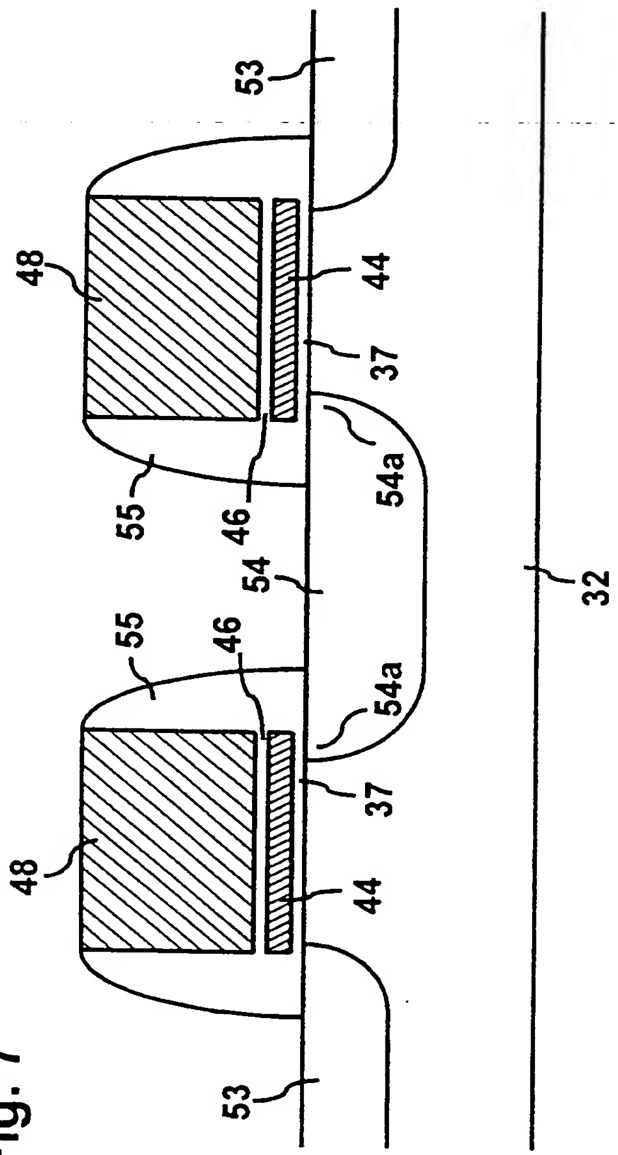


Fig. 7

Fig. 8A

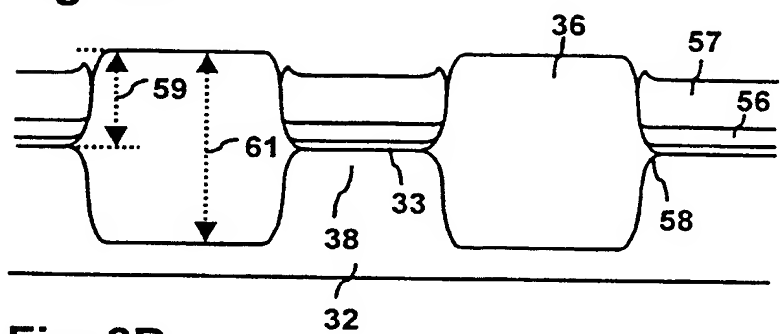


Fig. 8B

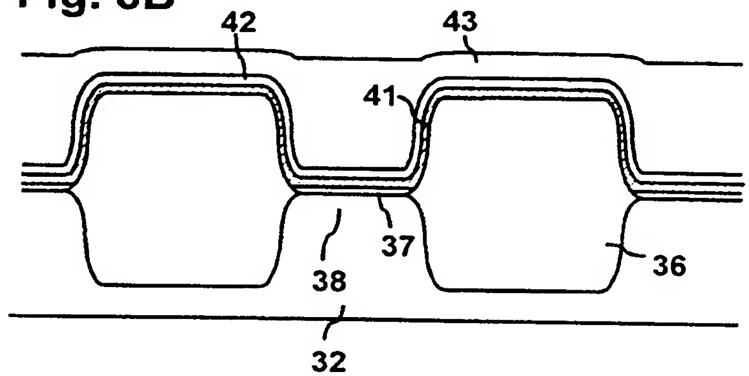


Fig. 8C

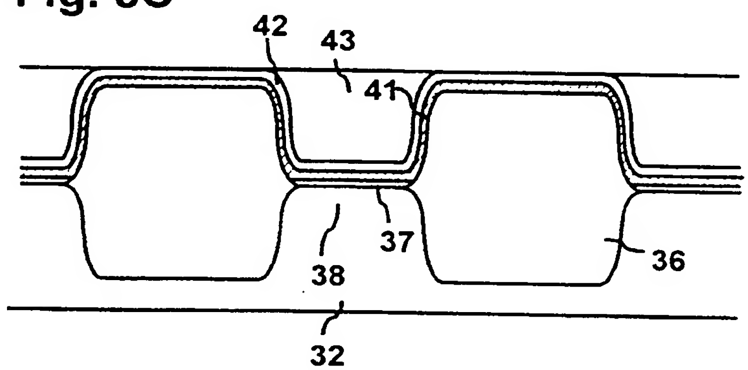


Fig. 8D

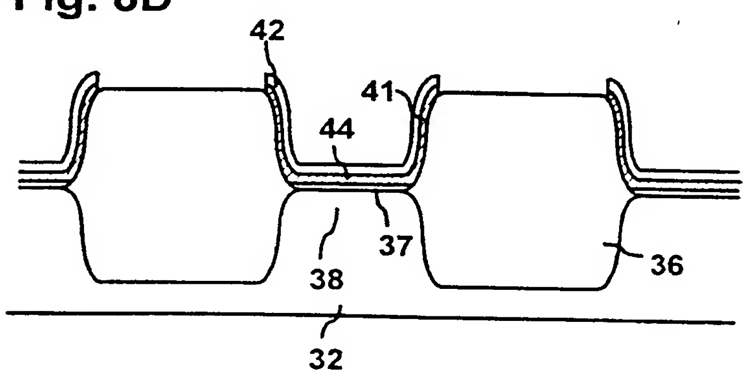


Fig. 8E

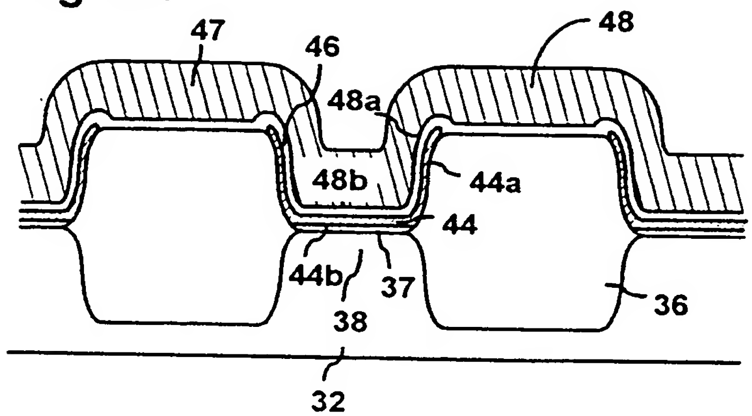
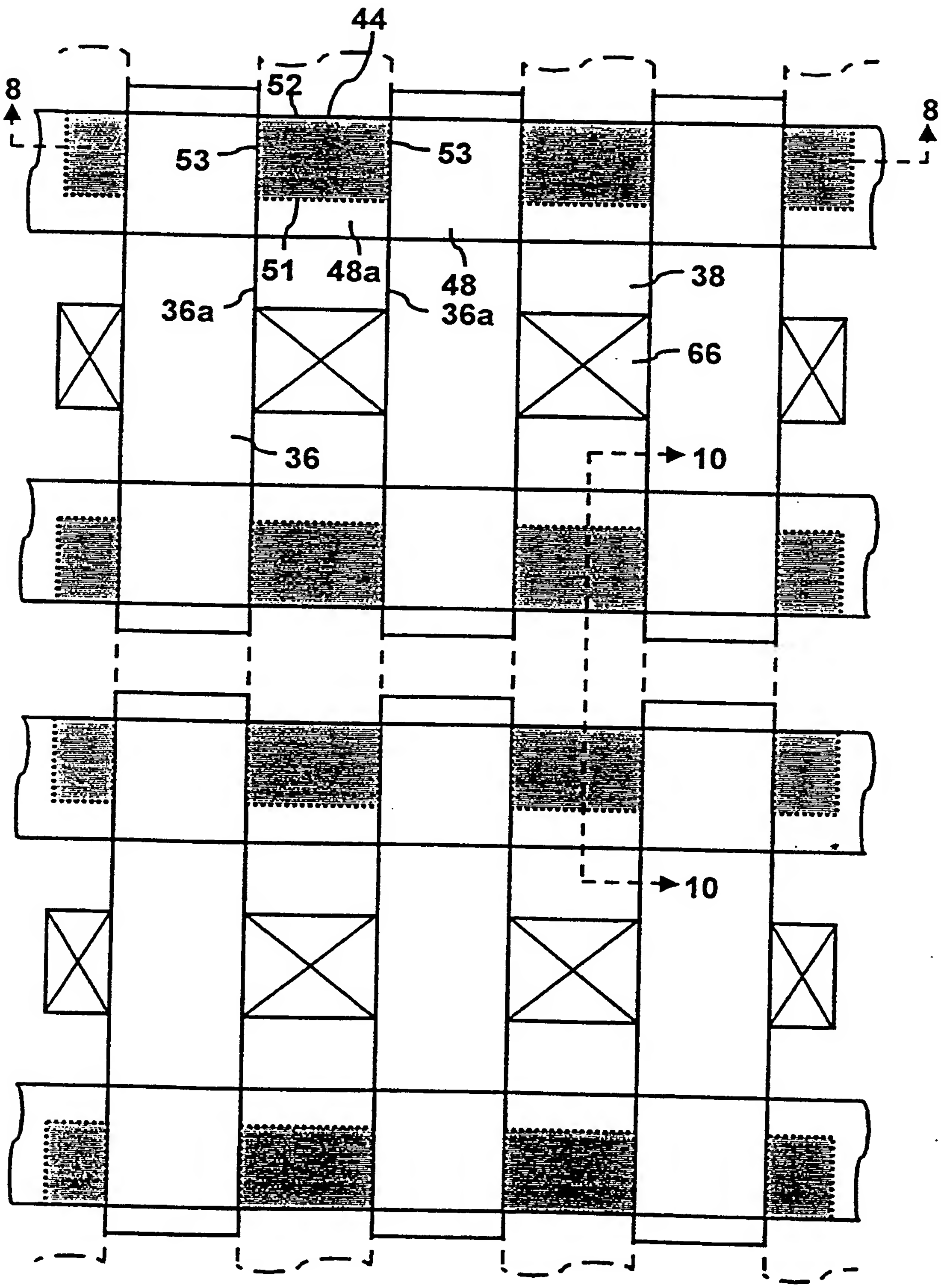


Fig. 9



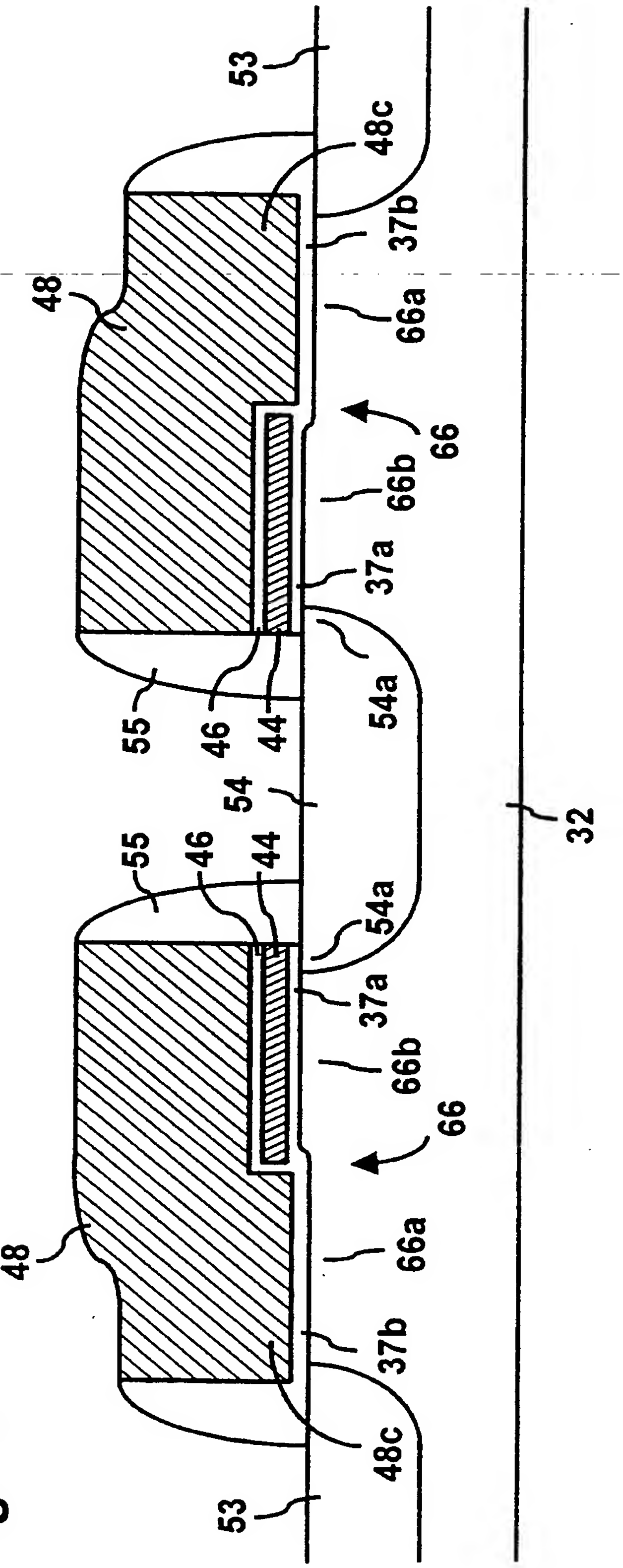


Fig. 10

Fig. 11A

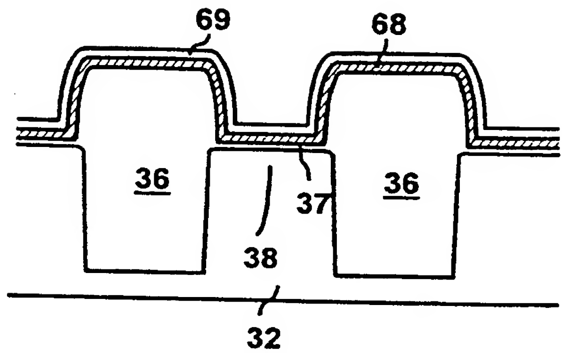


Fig. 11B

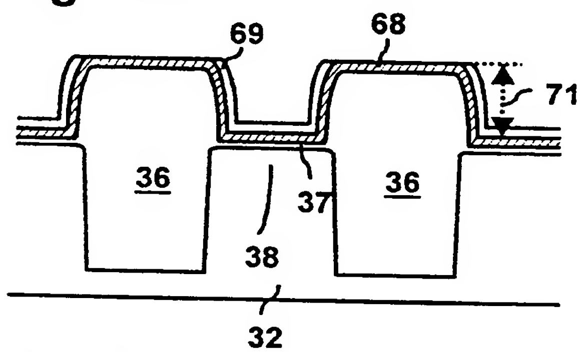


Fig. 12A

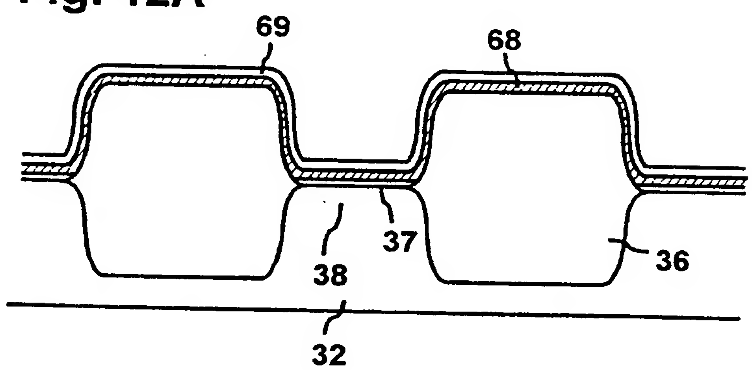
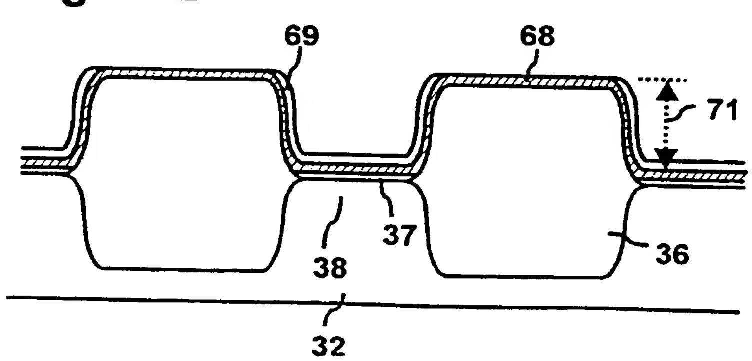


Fig. 12B



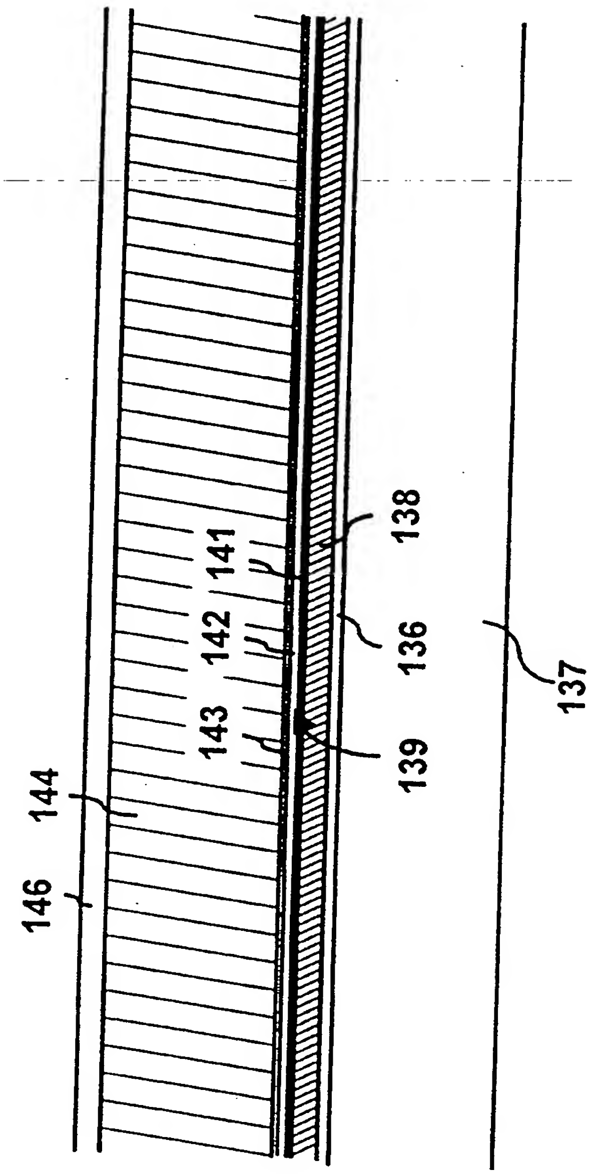


Fig. 13A

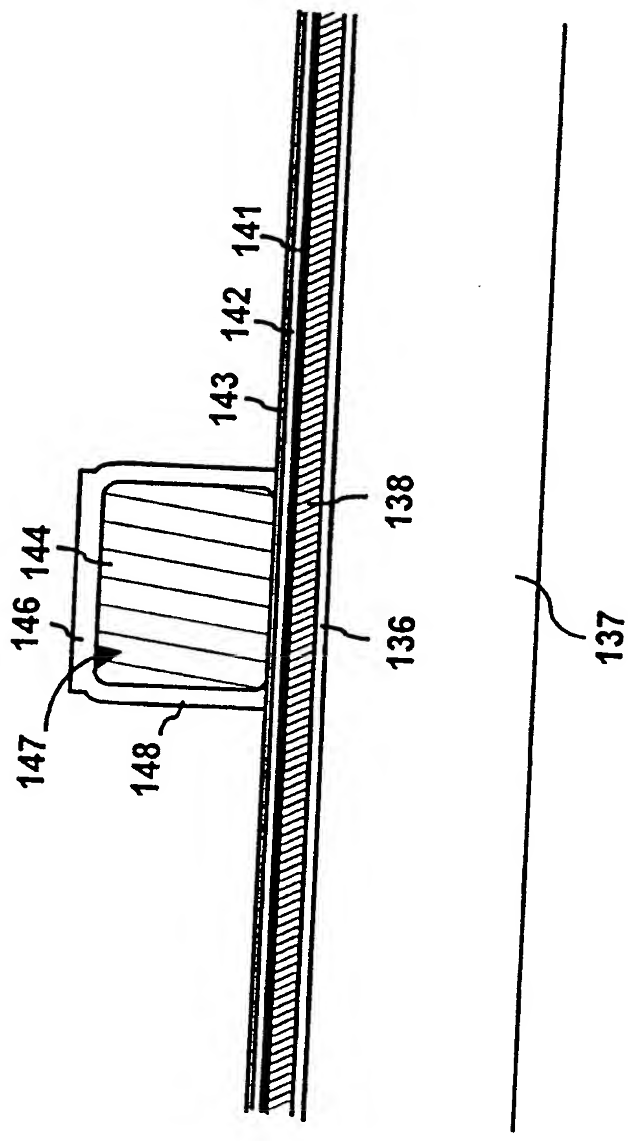


Fig. 13B

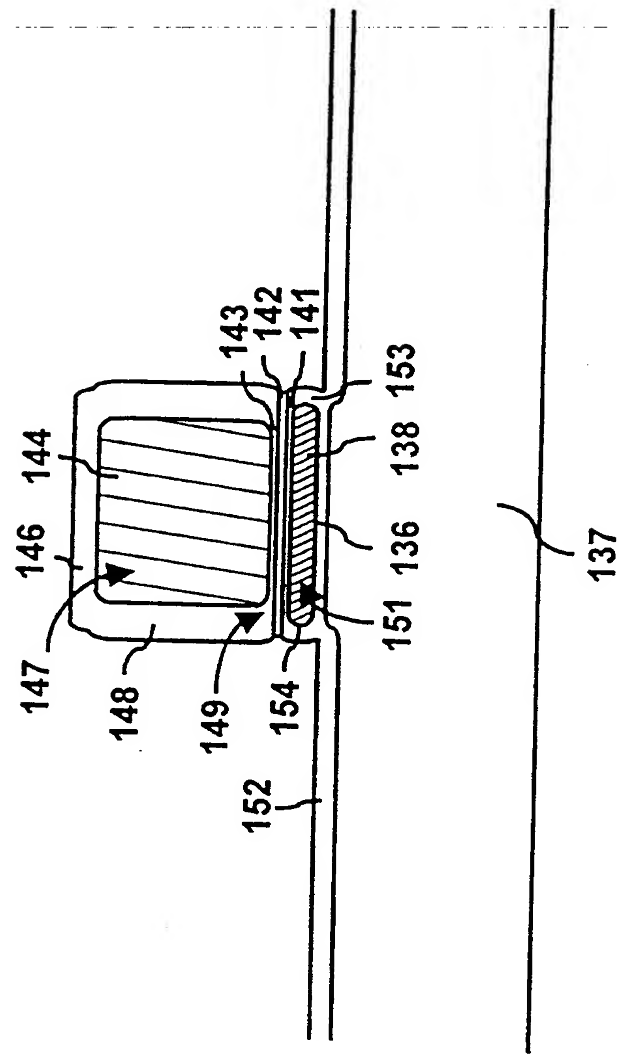


Fig. 13C

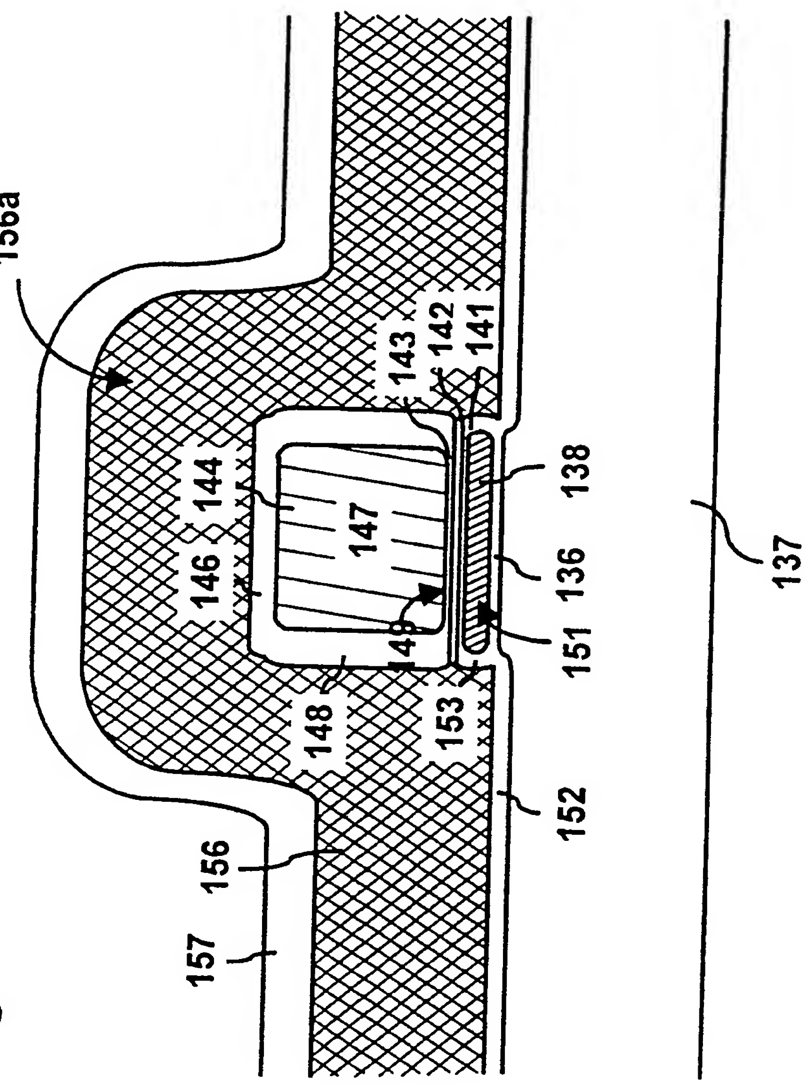

156a
143
142
141
144
146
147
138
136
137
148
151
148
153
152
156
157

Fig. 13D

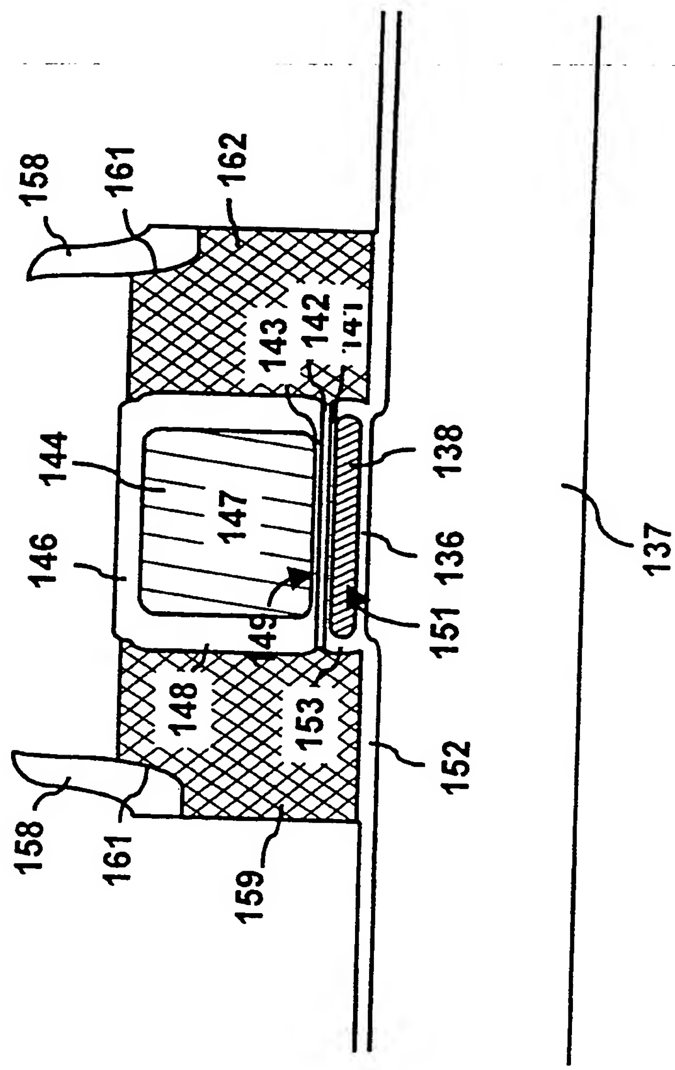
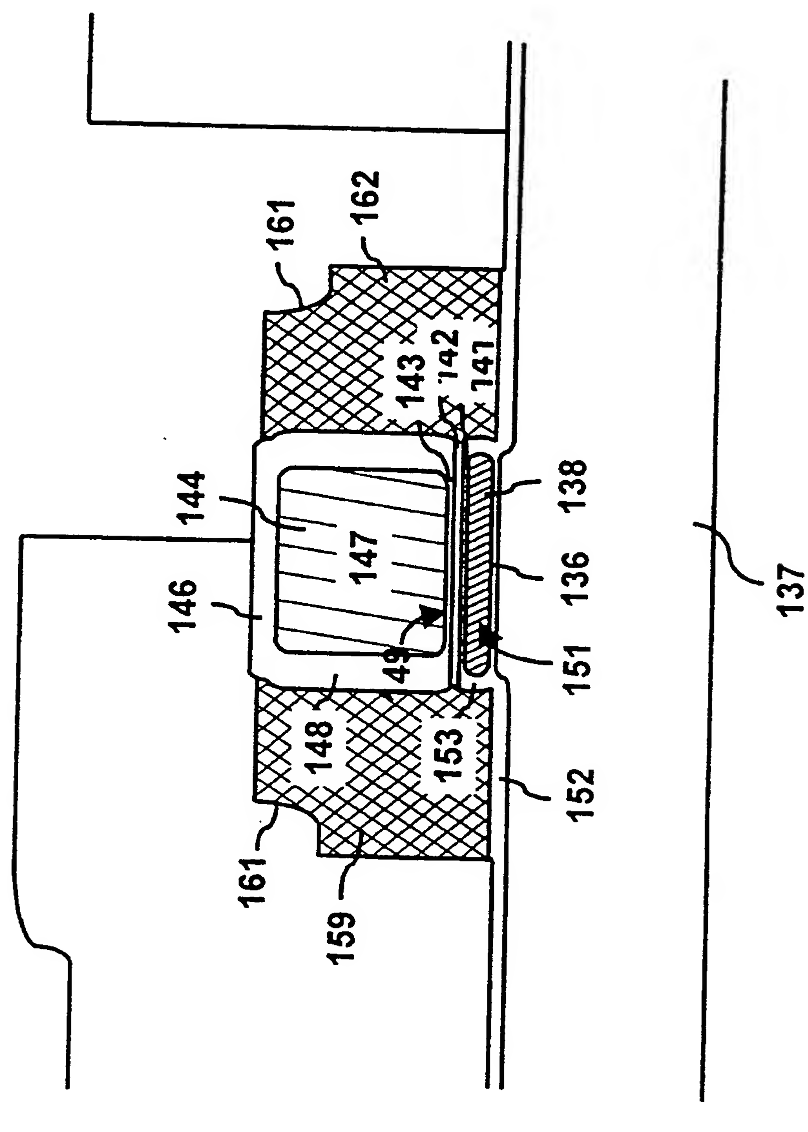

158
161
162
143
142
141
144
147
146
149
138
136
151
137
148
153
152
158
161
159

Fig. 13E

Fig. 13F


161
162
143
142
141
138
144
147
146
136
137
49
151
148
153
152
161
159

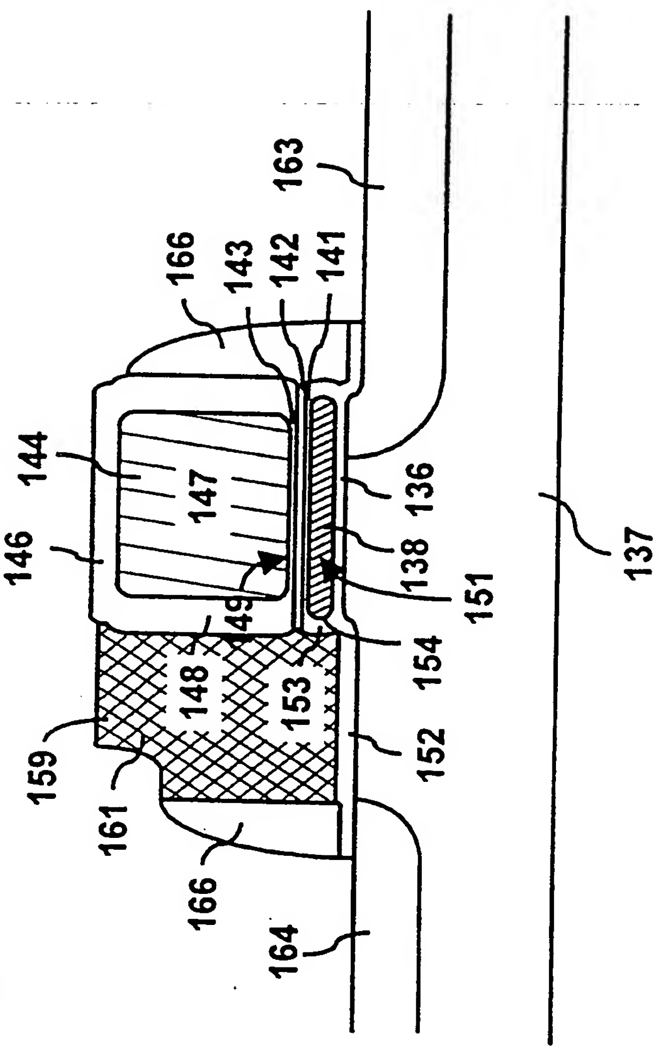

163
166
143
142
141
144
147
146
136
138
137
151
149
154
148
153
152
159
161
166
164

Fig. 13G

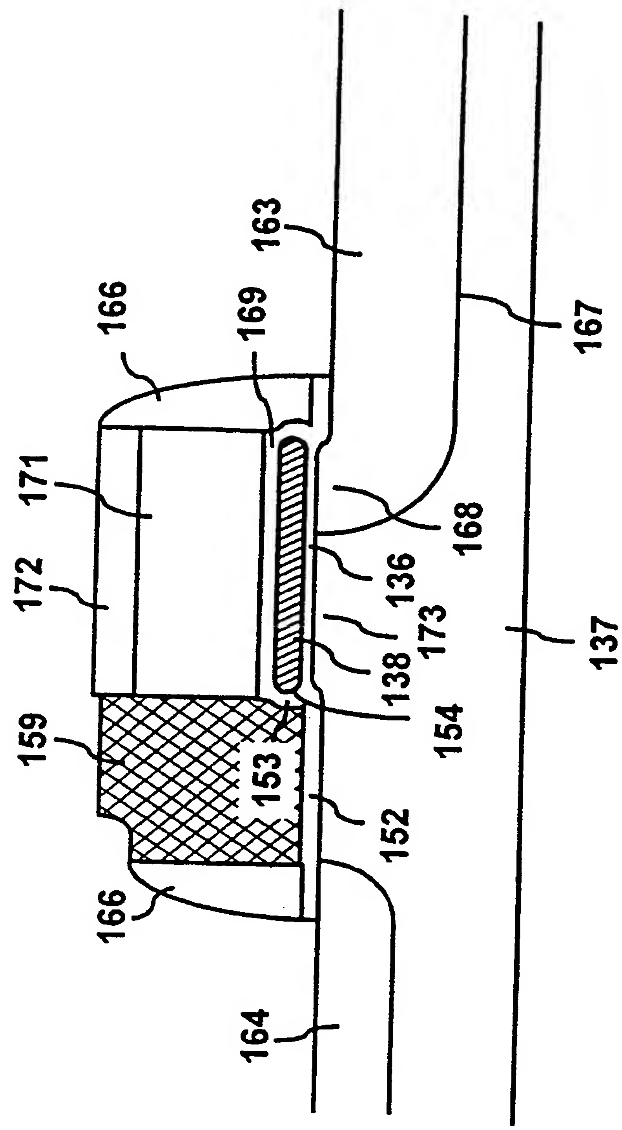

163
166
169
167
171
168
172
136
173
138
137
154
159
153
152
166
164

Fig. 14

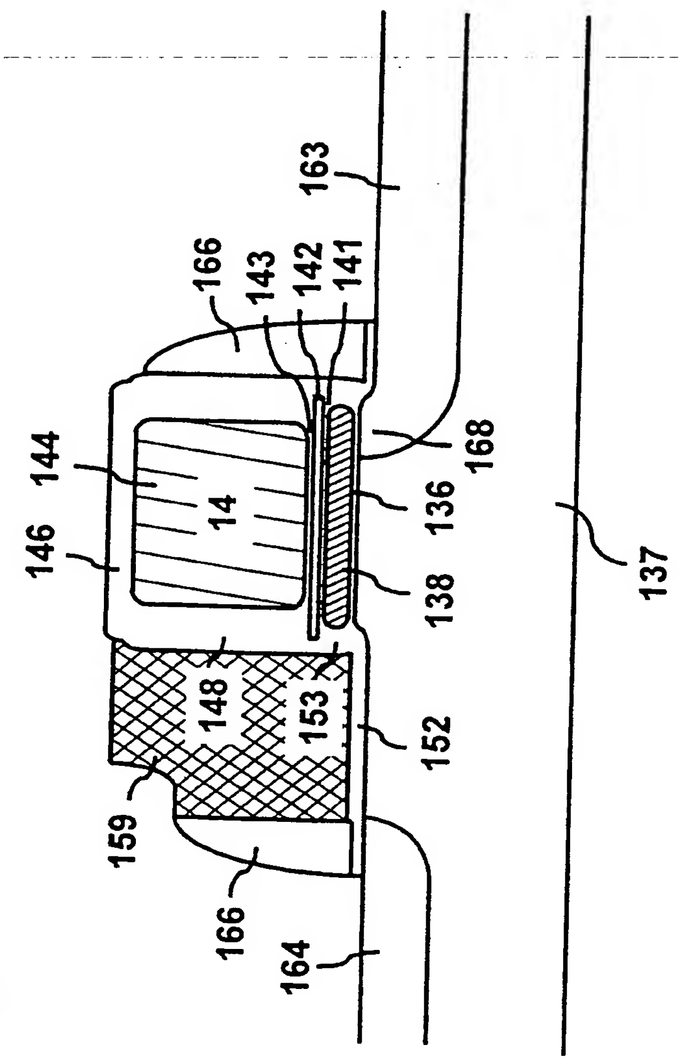

163
166
143
142
141
144
14
146
168
136
137
138
148
153
152
159
166
164

Fig. 15

Fig. 16

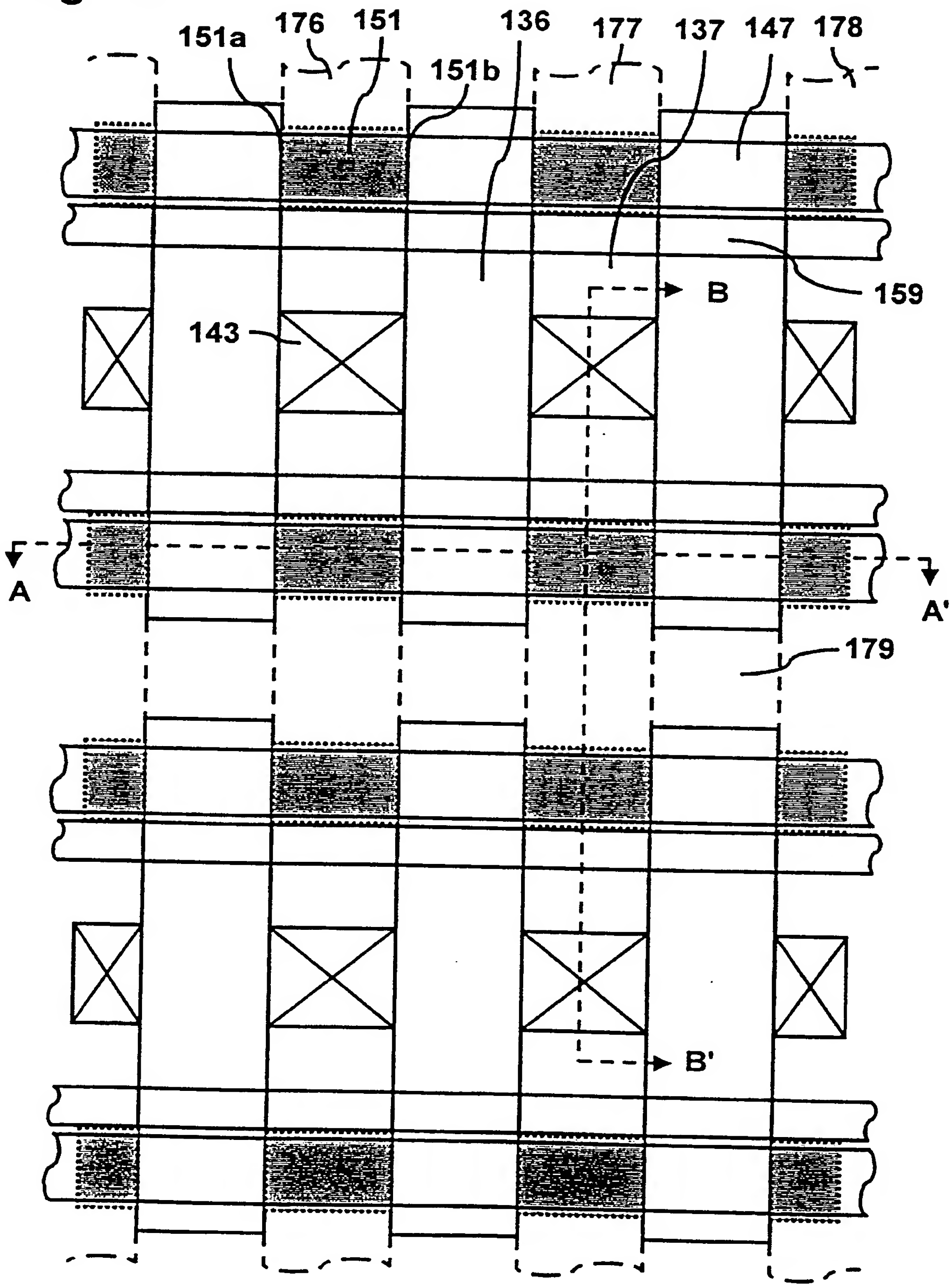


Fig. 17

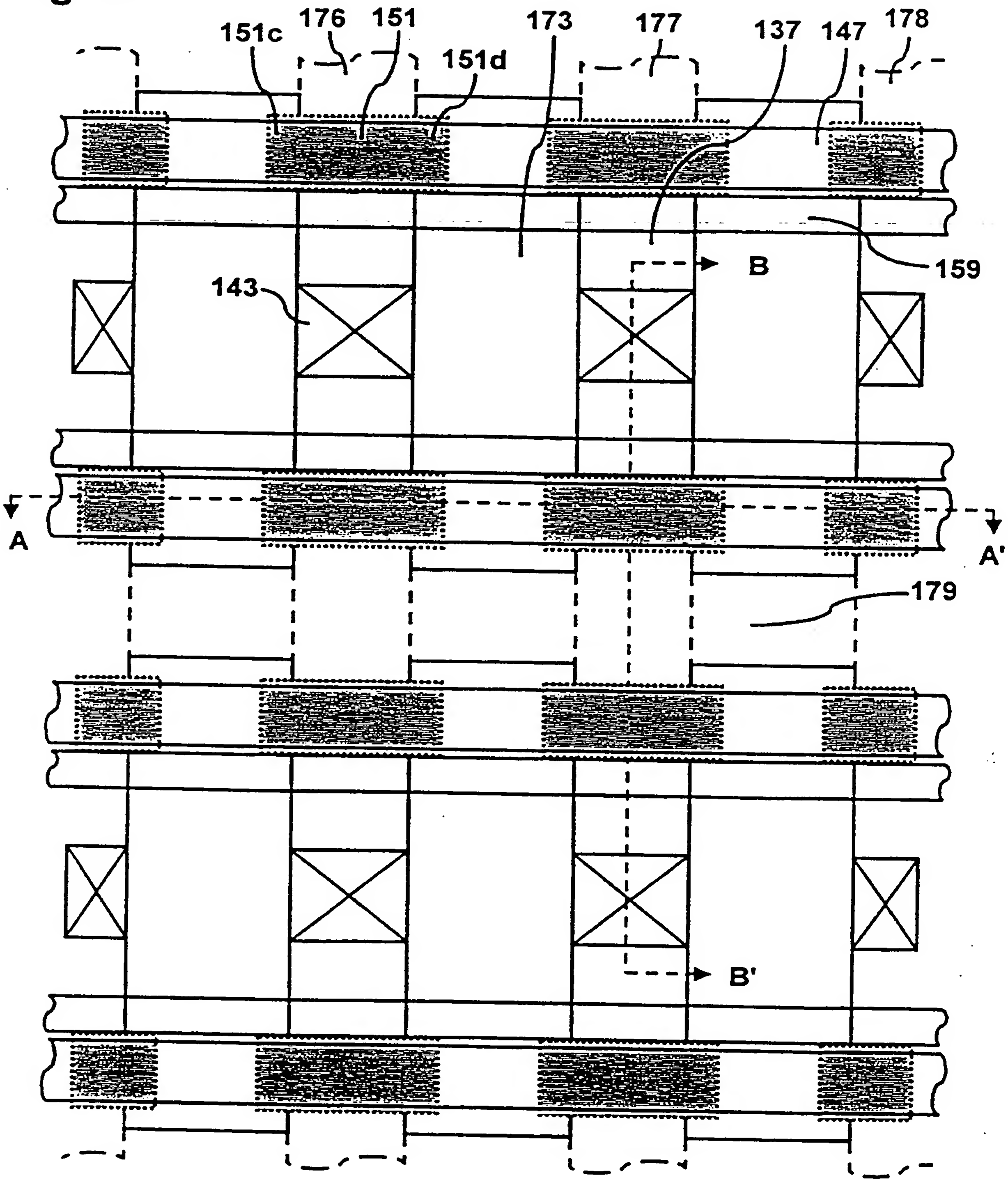


Fig. 18

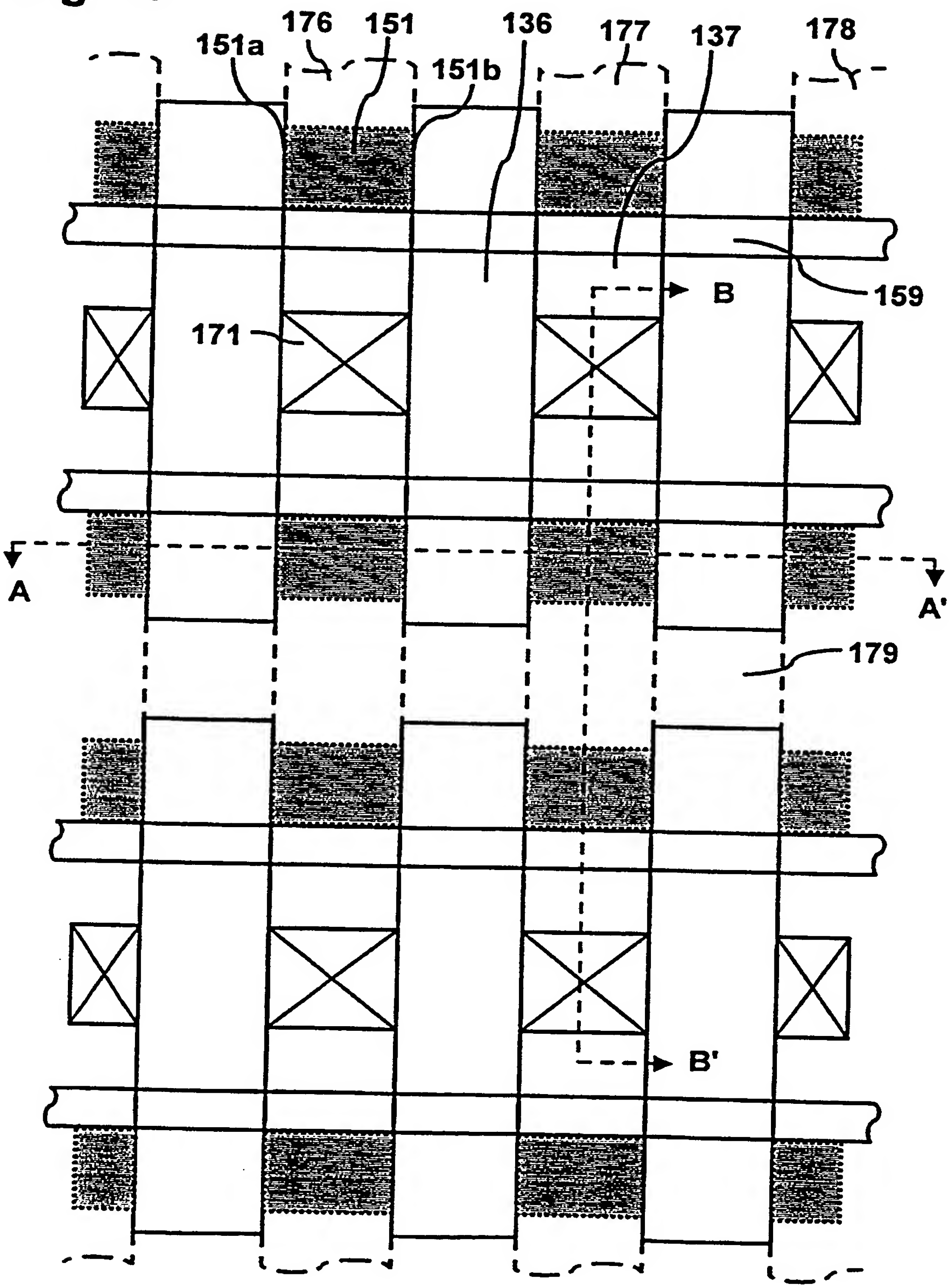


Fig. 19

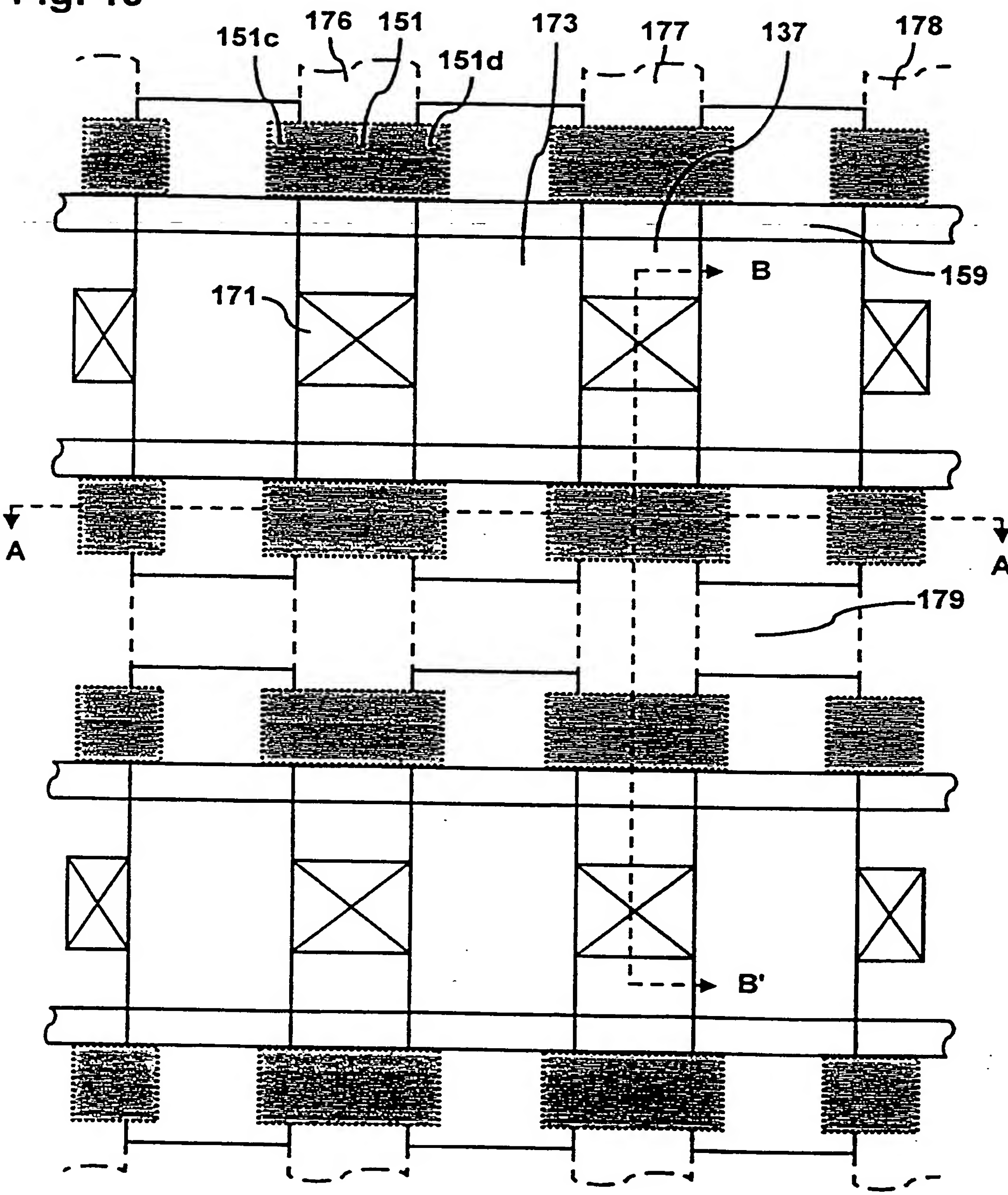


Fig. 20

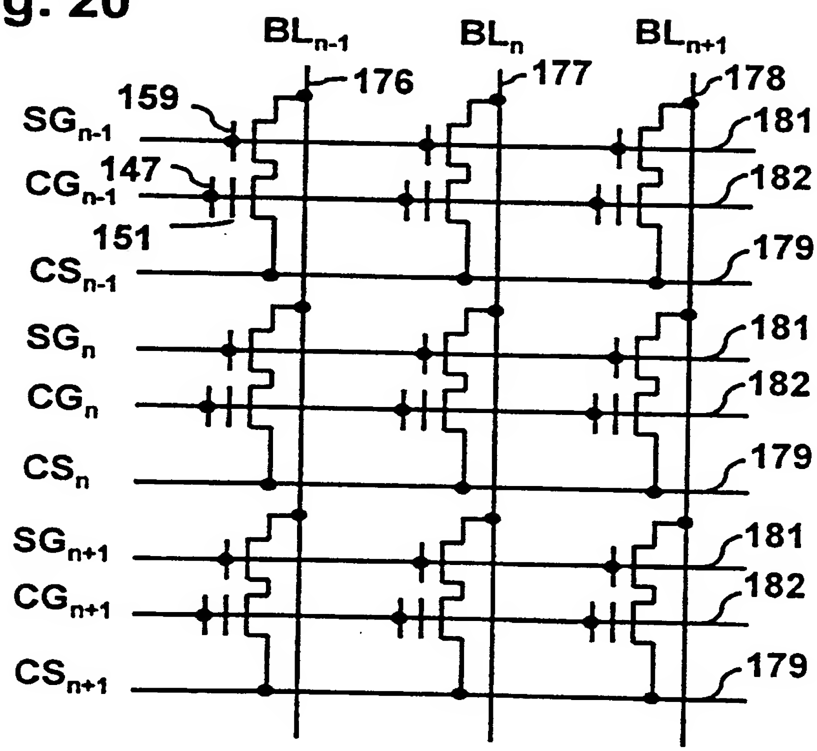
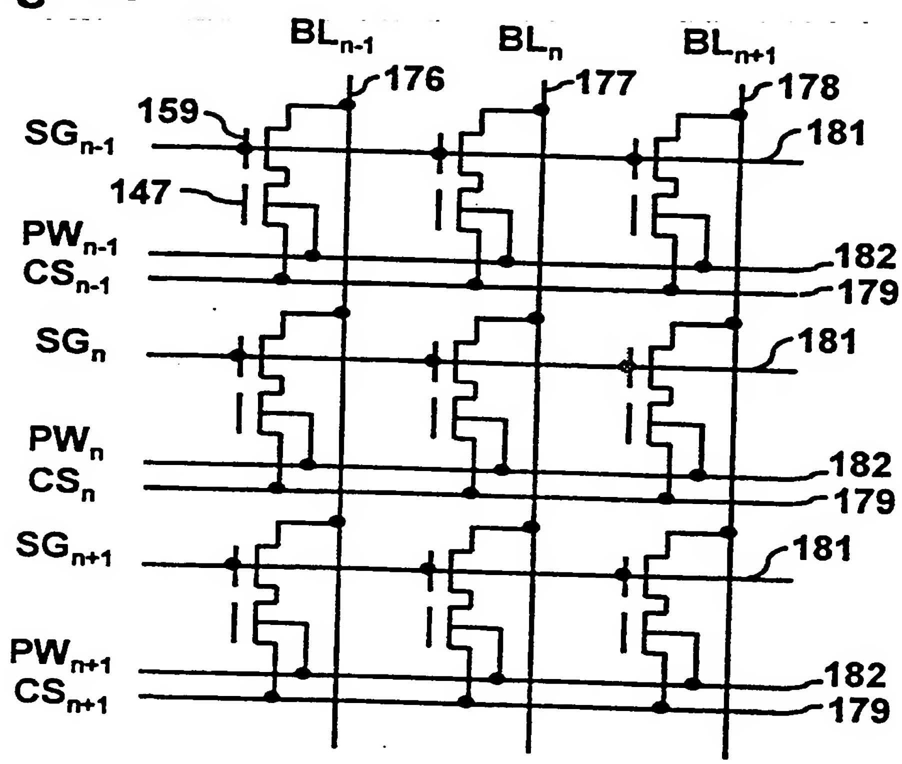

BL_{n-1}
BL_n
BL_{n+1}
176
177
178
159
SG_{n-1}
147
CG_{n-1}
151
CS_{n-1}
SG_n
CG_n
CS_n
SG_{n+1}
CG_{n+1}
CS_{n+1}
181
182
179

Fig. 21


BLn-1
BLn
BLn+1
176
177
178
159
181
SGn-1
147
PWn-1
CSn-1
182
179
SGn
181
PWn
CSn
182
179
SGn+1
181
PWn+1
CSn+1
182
179

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04455

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :H01L 29/788, 21/336
US CL : 257/315, 316, 317, 318, 321; 438/257, 259, 263, 264
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 257/315, 316, 317, 318, 321; 438/257, 259, 263, 264

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	U.S. 5,770,501 A (HONG) 23 June 1998 (23.06.98), col. 2, lines 55-63, col. 3, lines 10-51.	1-44

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 MAY 2000	13 JUN 2000

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	TOM THOMAS
Facsimile No. (703) 305-3230	Telephone No. (703) 308-2772